



### Yinshuang Xiao

Walker Department of Mechanical Engineering,  
 The University of Texas at Austin,  
 Austin, TX 78712-1591  
 e-mail: yinshuangxiao@utexas.edu

### Yaxin Cui

Department of Mechanical Engineering,  
 Northwestern University,  
 Evanston, IL 60208  
 e-mail: yaxincui2023@u.northwestern.edu

### Johan Koskinen

Department of Statistics,  
 Stockholm University,  
 Stockholm SE-106 91, Sweden  
 e-mail: johan.koskinen@stat.su.se

### Noshir Contractor

Department of Industry Engineering and  
 Management Science,  
 Northwestern University,  
 Evanston, IL 60208  
 e-mail: nosh@northwestern.edu

### Wei Chen

Department of Mechanical Engineering,  
 Northwestern University,  
 Evanston, IL 60208  
 e-mail: weichen@northwestern.edu

### Zhenghui Sha<sup>1</sup>

Walker Department of Mechanical Engineering,  
 The University of Texas at Austin,  
 Austin, TX 78712-1591  
 e-mail: zsha@austin.utexas.edu

# Product Design Incorporating Competition Relations: A Network-Based Design Framework Considering Local Dependencies

*System design has been facing the challenges of incorporating complex dependencies between individual entities into design formulations. For example, while the decision-based design framework successfully integrated customer preference modeling into optimal design, the problem was formulated from a single entity's perspective, and the competition between multiple enterprises was not considered in the formulation. Network science has offered several solutions for studying interdependencies in various system contexts. However, efforts have primarily focused on analysis (i.e., the forward problem). The inverse problem still remains: How can we achieve the desired system-level performance by promoting the formation of targeted relations among local entities? In this study, we answer this question by developing a network-based design framework. This framework uses network representations to characterize and capture dependencies and relations between individual entities in complex systems and integrate these representations into design formulations to find optimal decisions for the desired performance of a system. To demonstrate its utility, we applied this framework to the design for market systems with a case study on vacuum cleaners. The objective is to increase the sales of a vacuum cleaner or its market share by optimizing its design attributes, such as suction power and weight, with the consideration of market competition relations, such as inter-brand triadic competition involving three products from different brands. We solve this problem by integrating an exponential random graph model (ERGM) with a genetic algorithm. The results indicate that the new designs, which consider market competition, can effectively increase the purchase frequency of specific vacuum cleaner models and the proposed network-based design method outperforms traditional design optimization. [DOI: 10.1115/1.4066426]*

*Keywords:* system design, complex networks, optimization, design for market systems

## 1 Background and Introduction

In today's rapidly changing technological landscape, the increasing scale of autonomous agents in a system and their complex interactions are posing challenges for designers to effectively model system behaviors and optimize system performance [1]. To find effective solutions to these challenges, the literature has explored various methodologies. One notable work is the decision-based design framework, which models design as a decision-making process seeking to maximize the *value* of a designed artifact [2,3] by integrating customer preferences. Other significant contributions

include the application of Taguchi methods for robust system design [4], game-theoretic approaches to system optimization [5], and the utilization of Monte Carlo simulation for system modeling and design optimization [6]. However, one limitation in many of these works is the lack of effective ways to model dependencies between individual entities.

The recent development in network science has made it a suitable theoretical foundation to address this limitation, thus supporting research on various complex relationships between entities [7,8]. Some representative network-based design studies include (1) mechanical system representations in networks [9–11], (2) social network analysis in design, involving collaboration within virtual design communities [12–15] and co-evolution between design teams and product development [16–20], (3) ecological network-inspired design of engineered products, systems, and industries [21–25], (4) network models for design ideation [26,27], and (5) network-based customer preference modeling [28–32].

<sup>1</sup>Corresponding author.

Contributed by the Design Automation Committee of ASME for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received March 31, 2024; final manuscript received August 22, 2024; published online September 23, 2024. Assoc. Editor: Jie Zhang.

Furthermore, there has been growing interest in understanding the influence of local-level networks (as defined in Fig. 1) on system-level functionalities. Research efforts in this area include the use of network motif theory to study system resilience and reliability by assessing specific subsystem structures [33–37]. Other studies investigate how local-level network topologies (e.g., closed or open triangles) influence customer behaviors based on exponential random graph models (ERGMs) [38,39]. More recently, with the increasing popularity of graph neural network (GNN) models, their applications in complex systems engineering have been actively explored [40–44].

Despite the rapid development and wide application of network theories and models in the literature on systems engineering and design, particularly the emerging interest in studying the role of local networks in system design-related issues, the current focus remains on modeling and analyzing global- and local-level system information and their correlations to system performance. There is a scarcity of research on how the results obtained from network modeling and analysis, particularly the insights into significant local dependencies and global–local interactions, can be used to inform design decisions. The latter is an inverse problem. To solve this problem, there are multiple challenges. For example, (1) how do we identify essential local-level dependencies? (2) How do we formulate a system design problem considering those local-level dependencies? and (3) How to search for the optimal design in a network-transformed design space?

To address these challenges and advance the current state-of-the-art, we introduce a network-based system design framework that considers local dependencies. This framework includes: (1) mining significant local networks using network motif theory, (2) transforming the original system design objective using the identified local networks, and (3) searching optimal design attributes for desired system performance combining the genetic algorithm (GA) and an ERGM model. To demonstrate the utility of this framework, we conducted a case study using customer survey datasets on US household vacuum cleaners collected in our previous work [45]. Based on the proposed network-based design framework, this study makes two unique contributions: (1) we propose a novel network-based product design representation that incorporates local triadic competitions of individual products and (2) we develop a new algorithm for calculating the probability that target local triadic competitions exist and successfully integrate it into metaheuristic optimization to support the search for optimal design configurations.

The rest of the paper is organized as follows. In Sec. 2, we introduce the proposed design framework. Then, Sec. 3 demonstrates the design framework using the vacuum cleaner case study. In Sec. 4, we discuss the generalizability and limitations of the proposed framework. Finally, the paper is concluded in Sec. 5 with closing thoughts on our future work.

## 2 Methodology

In Fig. 2, we compare a typical system design process with our proposed network-based approach. The traditional method begins by analyzing the system requirements from which the design goal is set and the design variables and constraints are identified [46,47]. The design goal guides the formulation of the design objective function and the design variables and constraints define the design space. Subsequently, optimal/suboptimal design solutions are found by exploring and exploiting the design space by evaluating design candidates with the objective function. Finally, the design solution is validated against system-level requirements. Building upon the traditional framework, the proposed network-based design framework consists of six major steps, each of which is elaborated below.

**2.1 Step 1: Network Modeling.** The primary objective of the first step is to create a network representation, labeled as  $Y(X)$ , in

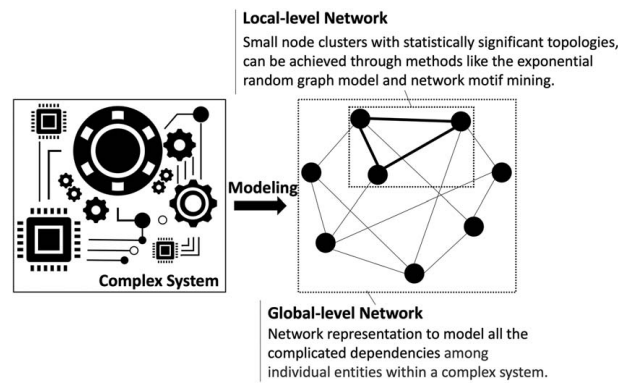


Fig. 1 Definitions of the global-level and local-level networks

which  $Y$  corresponds to the network’s adjacency matrix and  $X$  represents the vector of system design attributes.  $Y$  changes with  $X$ . Let us take the customer-product market system as an example, in which the co-consideration relations among products can be represented by a unidimensional network  $Y$  shown in Fig. 3. This network is built using data from customers’ considerations of vacuum cleaners and following the approach described in Ref. [45]. In such a network, each node represents a unique product model that the customers considered. The dashed links denote that two products from different brands (e.g., Dyson versus iRobot) are co-considered by at least one customer. In contrast, the solid links denote co-consideration within the same brand. We assume that the design attribute vector  $X = [x_1, x_2]$  in this example only includes the suction power and weight of each product. Updating the design attribute vector  $X$  for any product will influence the co-consideration and therefore the network structure  $Y$ .

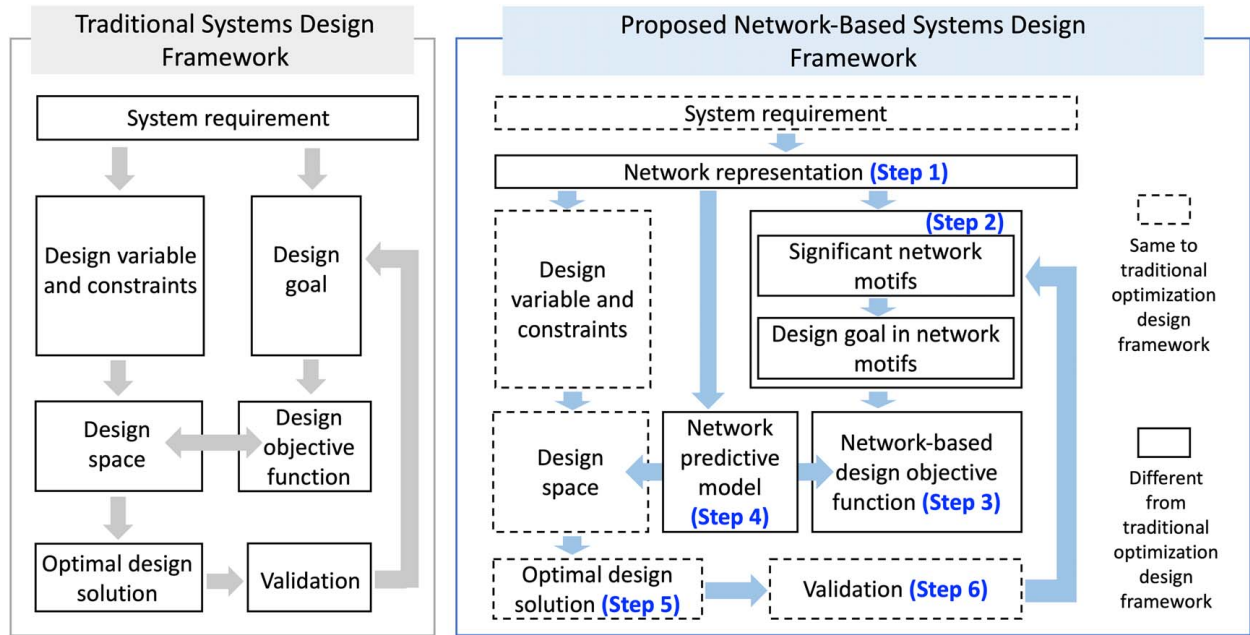
**2.2 Step 2: Representing the Design Goal Using Network Motifs.** The objective of Step 2 is to represent the design goal using local networks. This involves transforming the original design objective function  $u(X)$  into a function of local networks, denoted as  $u(g(y(X)))$ , where  $g(y(X))$  indicates the *derived local network-based design variable* (either a scalar or a vector). This step is essential to incorporate the significant dependencies between individual entities (represented by local networks) into the design process. To achieve the objective, we first identify significant local networks  $y(X)$  based on network motif theory [48].

Network motifs are the underlying non-random subgraphs within complex networks. Significant network motifs are found by comparing the real-world network  $Y_{\text{obs}}$  with a specific randomized network set<sup>2</sup>  $\Omega(Y')$  that includes  $N$  random networks. Only statistically significant local networks are considered as network motifs. The null hypothesis is that the frequencies of a local network in random networks  $F_{\text{rand}}(y)$  are equal to or greater than that of the real-world network  $F_{\text{obs}}(y)$ . It is rejected if the  $p$ -value given in Eq. (1) is less than a level of significance (commonly 0.01 or 0.05) [49].

$$p(y) = \frac{1}{N} \sum_{j=1}^N \delta(F_{\text{rand}}(y) \geq F_{\text{obs}}(y)) \quad (1)$$

where  $\delta$  is the sign function equal to 1 when  $F_{\text{random}}(y) \geq F_{\text{obs}}(y)$ , and 0 otherwise. In addition, the  $Z$ -score is another measurement

<sup>2</sup>For a rigorous comparison, each node in the random networks has the same number of degrees as the corresponding real-world network. Moreover, the random networks used to calculate the significance of size- $n$  local networks are generated to keep the same number of occurrences of all size- $(n-1)$  local networks as in the real-world network [48].



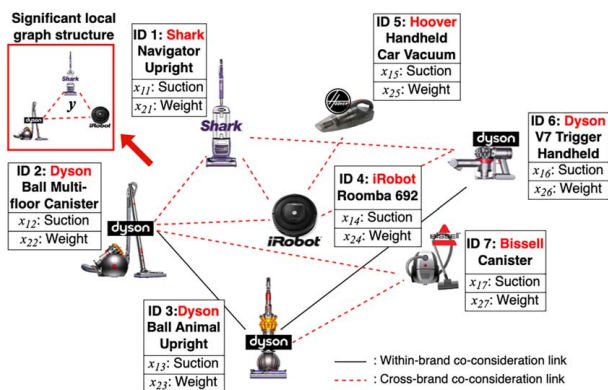
**Fig. 2 Comparison between the traditional system design framework and the proposed network-based system design framework considering local dependencies**

of the significance of a network motif, which is defined as

$$Z(y) = \frac{F_{\text{obs}}(y) - \mu_{\text{rand}}(y)}{\sigma_{\text{rand}}(y)} \quad (2)$$

where  $\mu_{\text{rand}}(y)$  and  $\sigma_{\text{rand}}(y)$  represent the mean and standard deviation of  $F_{\text{rand}}(y)$ . A higher  $Z$ -score indicates that  $y$  is a more significant motif in  $Y_{\text{obs}}$  [49].

Having identified the significant motifs, we proceed to convert the original design goal  $u(X)$  into the new representation in the form of local networks  $u(g(y(X)))$ , which capture interdependencies among individual entities (e.g., product models). To illustrate this, let us use an example of the vacuum cleaner market system. Suppose that we take on the role of Dyson, with the original design goal  $u(X)$  being to participate in the dominant product competition as much as possible. If a significant network motif  $y$  (shown in the top-left corner of Fig. 3), representing the inter-brand triadic closure competition, is identified as the significant competition pattern in this vacuum cleaner market, then we transform our goal into maximizing the number of inter-brand triadic closure competitions a product is involving in. Next, assume that the number is negatively correlated with the average suction power difference



**Fig. 3 An example of the customer-household vacuum cleaner market co-consideration network**

between products within the triadic closure. For a target product (e.g., product 2), the local network-based design variable  $g$  becomes a scalar  $g(y(X)) = (1/3)[|x_{12} - x_{11}| + |x_{12} - x_{14}| + |x_{11} - x_{14}|]$  and  $X = [x_{12}]$  (the values of  $x_{11}$  and  $x_{14}$  are given). The transformed design objective function represented by local networks is thus  $u(g(y(X))) = -(1/M(X)) \sum_{m=1}^{M(X)} g_m(y(X))$ . The negative sign indicates the assumed negative correlation between  $u$  and  $g$ .  $M$  is the total number of inter-brand triadic closure competitions in which product 2 is involved.

**2.3 Step 3: Optimization Problem Formulation.** In Step 3, an optimization problem is formulated based on the local network-based design objective function obtained in Step 2. Figure 4 illustrates an example of the optimization problem associated with the transformed design objective  $u(g(y(X)))$ . The objective function  $f(X)$  is defined to maximize the number of participation of all Dyson products in inter-brand triadic competitions,  $u_i$ , by adjusting the design attributes,  $x_{1i}$  and  $x_{2i}$ , for each product.  $i = 2, 3, 6$  stands for the product IDs of all Dyson products in Fig. 3. However, solving this network-transformed optimization problem is a challenge. This is because to obtain the number of triadic competitions  $M(X_i)$  in which the product  $i$  participates, we must know the network topology. But it changes every time when we change the design attributes (e.g., suction power stored as a node feature), and there is a lack of analytical expression between the design attributes and the network structure. Therefore, solving such an optimization problem that contains design variables in network representations (that could be non-linear) necessitates the employment of a surrogate model to predict new network structures when a node feature changes.

**2.4 Step 4: Predictive Model Training and Evaluation.** Step 4 is to establish a model to predict the new network topology after updating the design attributes  $X$ . Typical network models, such as ERGM [50] and GNNs [51], can be used for this purpose. In this study, we have chosen to adopt ERGM due to its verified performance in our previous work [38], and we will explore other models in our future work. ERGM is a stochastic network model for the observed network  $Y_{\text{obs}}$ , representing a particular instance within a collection of potential random networks  $Y$ . These

## Optimization Problem Example

$$\max f(\mathbf{X}) = \max [u_2 (g(\mathbf{y}(\mathbf{X}_2))) + u_3 (g(\mathbf{y}(\mathbf{X}_3))) + u_6 (g(\mathbf{y}(\mathbf{X}_6)))]$$

$$\text{S.t.} \quad \mathbf{X}_i = [x_{1i}, x_{2i}], i = 2, 3, 6$$

$$x_{suc\_low} \leq x_{1i} \leq x_{suc\_high}$$

$$x_{weig\_low} \leq x_{2i} \leq x_{weig\_high}$$

$u_i = -\frac{1}{M(\mathbf{X}_i)} \sum_{m=1}^{M(\mathbf{X}_i)} g_m(\mathbf{y}(\mathbf{X}_i))$ : the number of inter-brand triadic closure competitions that product  $i$  participates in, represented by the average suction difference among products within the triadic competitions. Note that the co-consideration network topology changes with product attributes. The number of triadic competitions  $M(\mathbf{X}_i)$  that product  $i$  engaged in is therefore the function of the attribute vector  $\mathbf{X}_i$ .

**Fig. 4 An example of optimization problem formulation with local network-based design objective**

random networks conform to the distribution described in Eq. (3) [38,50].

$$\Pr(Y = Y_{\text{obs}}) = \frac{\exp(\theta' \eta(Y_{\text{obs}}))}{\kappa(\theta)} \quad (3)$$

where  $\theta$  is a vector of model parameters, and  $\eta(Y_{\text{obs}})$  is a vector of the network statistics. These statistics can encompass either nodal attributes or edge attributes associated with design attributes  $\mathbf{X}$ . To ensure that Eq. (3) constitutes a valid probability distribution,  $\kappa(\theta)$  serves as a normalizing factor. It should be noted that the successful construction of the target ERGM is based on two factors. First, it is determined by the design attributes  $\mathbf{X}$ . This means that the attributes intended for the design must be included in the model through network statistics, such as nodal attribute effects [38]. Second, achieving a converged ERGM with satisfactory predictive power requires a trial-and-error process.<sup>3</sup> Three major categories of network statistics are given in Table 1.

ERGM training involves estimating the model parameters  $\theta$  by feeding the observed network data obtained from Step 1. Once the estimated parameters are obtained, the predictive performance of the estimated model can be evaluated in four steps. The first step is to use the estimated model to simulate  $N$  number of networks ( $Y_1, Y_2, \dots, Y_N$ ). Then, in the second step, each simulated network  $Y_n$  ( $n = 1, \dots, N$ ) is compared against the observed network  $Y_{\text{obs}}$  (that is, the ground truth) in Step 1 to classify all possible links of the network in a confusion matrix [55]. According to the confusion matrix, we can then calculate common metrics, such as *Recall*, *Precision*, and *F1-Score*, to evaluate the predictive performance of the model [55]. Finally, the mean values of *Recalls*, *Precisions*, and *F1-Scores* of those  $N$  simulated networks are used to represent the performance of a trained ERGM.

<sup>3</sup>An ERGM failing to converge indicates that the parameter estimates are not settling down to stable values, and the iterative estimation process is not reaching a consistent solution [52]. Typical reasons for the convergence issue of ERGM include model degeneracy (an ill-fitting model in ERGM fails to adequately represent the observed network) [53] and inappropriate model specifications [52].

**2.5 Steps 5 and 6: Optimal Problem Solving and Solution Validation.** In Step 5, the model obtained from Step 4 will be used as a surrogate model to predict new network structures in order to re-evaluate the objective value  $f(\mathbf{X})$  after modifying the associated node features (i.e., design attributes). Consequently, the computational search process for the optimal solution is performed through metaheuristic approaches, such as the genetic algorithm [56] or particle swarm optimization [57]. After finding the optimal design solution in Step 6, it can be validated by implementing the new designs and observing if the desired system performance can be achieved or not. For example, in the vacuum cleaner example, we will count if the new design attributes of a particular product would help increase its participation in the desired competition relations on the market. Validation is often challenging as it requires real-world implementation and testing. In this paper, we focus primarily on the verification of the optimization results computationally by recalculating the design objective with the optimal design variables and checking if the objective value indeed increases or not.

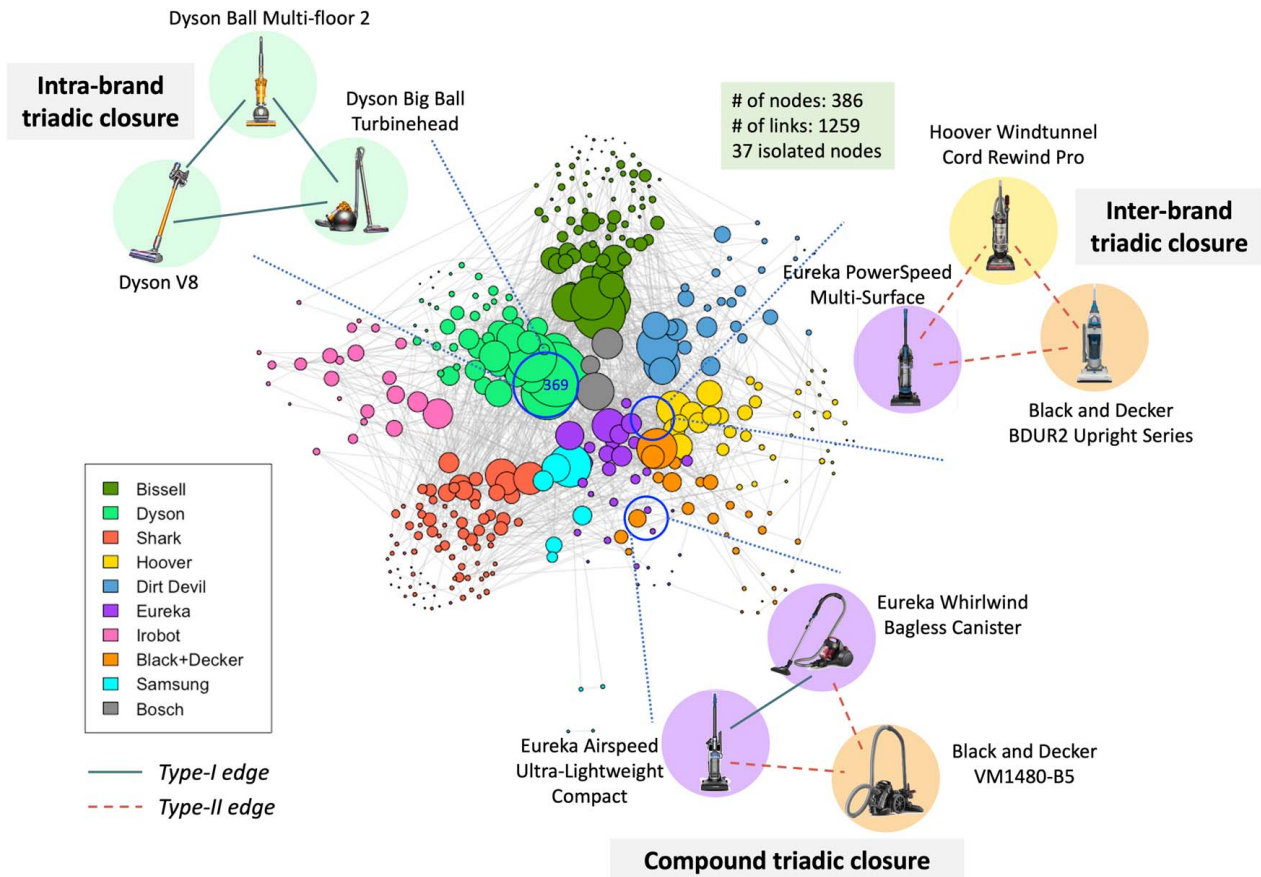
## 3 Case Study

In this section, the US household vacuum cleaner market system is used as a case study to demonstrate the proposed network-based system design framework.

**3.1 Data Source and Network Modeling.** *Data Source:* The dataset used in this study is obtained from our previous survey study [45] launched in 2021 on Cint, a provider of digital survey solutions. The publicly available dataset contains 1002 responses from vacuum cleaner buyers, covering 624 unique models of household vacuum cleaners [58]. The dataset covers a diverse array of attributes, spanning from customer demographics and social

**Table 1 Three major network statistics in ERGM with their examples [38,54]**

Category	Examples	Interpretation
Nodal attributes effects	<i>Nodecov</i>	Main effect of a covariate. For example, <i>nodecov.suction</i> understands how suction power influences a vacuum cleaner being co-considered with other vacuum cleaners.
Relational attributes effects	<i>Absdiff</i>	Absolute difference between two connected nodes' attributes. For example, <i>absdiff.weight</i> looks into whether a large or small weight difference between two vacuum cleaners motivates them to be co-considered.
Network structural effects	<i>Edges</i>	Equal to the number of links in the network, equivalent to the intercept term in the regression model. In the context of the vacuum cleaner co-consideration network, it estimates the likelihood that two vacuum cleaners will be co-considered randomly.
	<i>GWESP</i>	Geometrically weighted edgewise shared partner. In Ref. [53], it is also called $k$ -triangle, which is defined to be a set of $k$ distinct triangles that share a common edge. The <i>GWESP</i> term models the tendency for edges that close triangles to be more probable than edges that do not close triangles. In the context of the vacuum cleaner co-consideration network, it investigates whether two vacuum cleaners co-considered with the same set of vacuum cleaners are more likely to be co-considered or not.



**Fig. 5 Co-consideration network of top ten household vacuum cleaner brands with a specific example for each significant competition motif. The *type-I edge* indicates that two vacuum cleaners share the same brand, and *type-II edge* refers to the different brands.**

network information to technical specifications of vacuum cleaners. The participants were tasked with detailing both the options that they initially considered and the ones they ultimately purchased.

**Co-consideration Network Model:** In this study, we are interested in competition analysis in the vacuum cleaner market system. We, therefore, construct the co-consideration network following our previous study [59]. In this unidimensional network, the nodes are unique vacuum cleaners from the top ten dominant brands and are considered by customers. Similar to the example given in Fig. 3, the undirected links represent that two vacuum cleaners are co-considered by at least one customer. The visualization of the co-consideration network is shown in Fig. 5. This network

contains 386 unique vacuum cleaner models and 1259 co-consideration links. Product 369, Dyson Ball Multi-floor 2, has the largest degree, indicating that it is co-considered most frequently.

### 3.2 Deriving the Local Network-Based Design Goal and Formulating the Optimization Design Problem.

**Definition of Derived Local Network-Based Design Variable:** As described in Sec. 2.2, the first step of deriving the local network-based design goal is to identify significant local network structures [59]. As shown in Table 2, three significant network motifs of the co-consideration network are identified by the motif mining tool, FANMOD [60], each of which represents distinct competition relationships between brands (inter-brand) and within a brand (intra-brand). They are named for their edge types and topological characteristics. A real-world example for each significant motif is also given in Fig. 5. Among the three motifs, the inter-brand triadic closure competition with the highest Z-score is found to be the most significant competition structure in the co-consideration network. Based on the positional characteristics in these significant motifs, we can define four unique node roles:  $R_1$ ,  $R_2$ ,  $R_3$ , and  $R_4$ . For example, in the inter-brand triadic closure, all three node positions share the same type of node role,  $R_1$ , because each node is co-considered with two products from two other brands in a closed triangle competition. Accordingly, we define the derived local network-based design variable of each product as  $g(y(X)) = [N_{R1}, N_{R2}, N_{R3}, N_{R4}]$ , where  $N_{Ri}$  (for  $i = 1, 2, 3, 4$ ) is the number of times a product is involved in the node role  $R_i$ . This defined network-based design variable can be easily extended. For example, if additional node roles such as  $R_5$  and  $R_6$  are discovered, we can extend the derived variable by concatenating  $N_{R5}$  and

**Table 2 Significant size-3 competition motifs in the co-consideration network**

Motif	Z-score
Inter-brand triadic closure competition	317.13
Intra-brand triadic closure competition	263.14
Compound triadic closure competition	86.65

Type-I edge	Type-II edge	Unique node role

**Table 3 Negative binomial regression estimated result of  $u$  corresponding to  $g(y(X)) = [N_{R1}]$**

Independent variables	Est. Coeff.	Std. error
Intercept	0.316***	0.075
$N_{R1}$	0.117***	0.017
$N_{R1}^2$	-0.002***	0.0005

\*\*\*0.000 level of significance.

$N_{R6}$ , resulting in  $g(y(X)) = [N_{R1}, N_{R2}, N_{R3}, N_{R4}, N_{R5}, N_{R6}]$ . Similarly, the vector can be shortened by omitting less important node roles.

**Optimization Problem Formulation:** Now, let us pick one particular product model, e.g., Dyson Ball Multi-floor 2 (Product 369)—the one with the most co-consideration connections in the observed network—to continue the demonstration due to its increasing popularity in the US market. Assuming that Dyson is interested in maximizing a product's market share, we use the number of purchases as an indicator of that product's market share. Next, we formulate the network-based design objective function by estimating the relationships between the number of times product purchases  $u$  and the local network-based design variable derived  $g(y(X))$ . As shown in Table 2, given that inter-brand triadic closure shows the highest Z-score, we simplify the derived design variable by focusing only on the node role  $R1$  in our first test case, resulting in  $g(y(X)) = [N_{R1}]$ . In this study, since the data format of the number of times product purchases is a count, following the method introduced in Ref. [61], negative binomial regression is selected to estimate the relationship between  $u$  and  $g(y(X))$ . To ensure the reliability of the estimate model, four models corresponding to the combination of polynomial terms of the independent variable up to cubic are tested. Both the mean absolute error (MAE) [62] and Akaike's information criterion (AIC) [63] are applied to measure the goodness of fit of the model while considering a balance between the goodness of fit and model complexity. The model with the lowest AIC and MAE is finally selected, which is provided in Table 3.

Therefore, the design objective function for the derived local network-based design variable is given in Eq. (4).

$$u(g(y(X^{P369}))) = \exp(0.316 + 0.117N_{R1}^{P369} - 0.002(N_{R1}^{P369})^2) \quad (4)$$

where  $N_{R1}^{P369}$  is the number of times Product 369 is involved in node role  $R1$ . This is equivalent to the number of inter-brand triadic closure competitions that include Product 369. Thus, we can further express  $N_{R1}^{P369}$  as shown in Eq. (5).

$$N_{R1}^{P369}(y(X^{P369})) = \frac{1}{2} \sum_{m=1}^M \det(y_{\text{inter\_brand}}^m(X^{P369})) \quad (5)$$

In this equation,  $M$  represents the number of all potential inter-brand triadic closure competitions in which Product 369 is involved. These competitions can be enumerated when we know the total number of products on the market. Suppose that there are  $K$  products on the market, and we denote the product set as  $V$  where each element is a product model named by its ID (e.g., product 369's ID name is  $P369$ ). In that case, we can pre-save the inter-brand triadic closure competitions in which  $P369$  is involved as a set denoted as  $S = \{y_{P369, V_i, V_j}^m\}$  where  $V_i$  and  $V_j$  represent the  $i$ th and  $j$ th products in  $V$  with which Product 369 competes. For instance, in the example shown in Fig. 3,  $V$  denotes  $\{P1, P2, P3, P4, P5, P6, P7\}$ , so  $K = 7$ . Among these products,  $P1, P4, P5$ , and  $P7$  are the only products that can form the inter-brand triadic closure competition with product 2. As a result,  $M = 6$  for product 2, and  $S = \{y_{P2, P1, P4}^1, y_{P2, P1, P5}^2, y_{P2, P1, P7}^3, y_{P2, P4, P5}^4, y_{P2, P4, P7}^5, y_{P2, P5, P7}^6\}$ .

In the case study, according to the network model shown in Fig. 5, we get  $K = 385$  and  $M = 46,650$  for product 369.  $y_{\text{inter\_brand}}$  represents the adjacency matrix for the inter-brand

## Optimization Problem Formulation

Maximize the number of times Product 369 purchases.

$$\max u(g(y(X^{P369}))) = \max \exp(0.316 + 0.117N_{R1}^{P369}(y(X^{P369})) - 0.002(N_{R1}^{P369}(y(X^{P369})))^2)$$

$N_{R1}^{P369}(y(X^{P369}))$ : the number of times Product 369 is involved in node role  $R1$ .

$$X^{P369} = [x_s^{P369}, x_w^{P369}]$$

$$\text{S.t.}^* \quad x_s^{P369} \in [x_{suc\_low}, x_{suc\_high}] = [1, 5]$$

$$x_w^{P369} \in [x_{weig\_low}, x_{weig\_high}] = [3.34, 29.3]$$

$$x_p^{P369} = \$284.98$$

\*: Given that Product 369 belongs to upright vacuum cleaner, we establish the design space's lower and upper bounds by utilizing the minimum and maximum suction power and weights observed among upright vacuum cleaners available in the market.

**Fig. 6 The optimization problem formulation**

triadic closure competitions. It equals to

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

when the triadic closure exists. Accordingly, the determinant of  $y_{\text{inter\_brand}}(X^{P369})$ , denoted as  $\det(y_{\text{inter\_brand}}(X^{P369}))$ , equals 2. The existence of the inter-brand triadic closure is determined by the original product design vector  $X^{P369}$ . In this case study, we focus on two essential design attributes of vacuum cleaners: suction power ( $x_s$ ) and weight ( $x_w$ ). Therefore,  $X^{P369} = [x_s^{P369}, x_w^{P369}]$ . In addition, we factor in one constrained variable: price ( $x_p$ ). The reason for having this constraint is that price is one of the most important factors influencing customer choice behaviors, and keeping it unchanged helps avoid the price perception bias that customers have toward the new design [64]. Finally, the formulated optimization problem is presented in Fig. 6.

**3.3 ERGM-Based Network Prediction.** As aforementioned, this study focuses on two design attributes and one constraint: suction power ( $x_s$ ), weight ( $x_w$ ), and price ( $x_p$ ). Therefore, they are taken into account by incorporating their associated *Nodecov* and *Absdiff* terms [50], as introduced in Table 1. To ensure model convergence, a trial-and-error process is performed. We evaluated 27 models with varying combinations of *Nodecov* and *Absdiff*, ultimately pinpointing a converged ERGM with all terms achieving a level of significance (*p-value*) close to 0. As presented in Table 4, in addition to three nodal effect terms, two network effect terms given in Table 1, *Edges* and the *GWESP*, are also included. To facilitate the convergence of the model and improve its performance, max-min normalization is applied to preprocess the attribute data [65]. The estimated results, i.e., the estimated model parameters  $\theta$  in Eq. (3), are shown in Table 4. For example, the negative sign of *Absdiff.price* shows that two vacuum cleaners with less difference in their prices are more likely to be co-considered. In contrast, the positive sign of *GWESP* means that two vacuum cleaners that share the same set of co-consideration products are more likely to be co-considered with each other. It implies that customer's consideration decisions involve a form of multiway grouping and comparison [39].

Once a trained model is obtained, we follow the process introduced in Sec. 2.4 to simulate 100 networks and validate its predictive power. By comparing the simulated networks against the

**Table 4 Estimated result of the predictive ERGM**

Independent variables	Est. Coeff.	Std. error
Edges/intercept	-6.717***	0.105
Absdiff. price	-0.562***	0.148
Absdiff. weight	-1.287***	0.194
Nodecov. suction	0.187***	0.035
GWESP	2.671***	0.092

\*\*\*0.000 level of significance.

**Table 5 Case 1: negative binomial regression estimated result of  $u$  corresponding to  $X = [x_s]$** 

Independent variables	Est. Coeff.	Std. error
Intercept	1.110***	0.310
$x_s$	-0.432	0.231
$x_s^2$	0.087*	0.039
$x_p$	0.0003	0.0003

\*\*\*0.000 level of significance.

\*0.01 level of significance.

·0.05 level of significance.

ground truth, we calculate the means of the *Precisions*, *Recall*, and *F1-Scores* of all 100 simulated networks, and the results are 0.158, 0.192, and 0.173, respectively. The *Precision* 0.158 indicates that 15.8% predicted co-consideration links are correctly predicted on average. The *Recall* 0.192 means that 19.2% truly existing links are correctly predicted. Lastly, *F1-Score*, the harmonic mean of *Precision* and *Recall*, evaluates a balanced predictive accuracy of the model. It should be noted that in this study we did not spend excessive resources to find the best ERGM model and only included four independent variables in the model, with the objective of this case study being to demonstrate the proposed design framework. Therefore, we focused on more model convergence and stopped testing additional models (which required more data collection efforts) for better prediction. A further discussion of this is presented in detail in Sec. 4.

**3.4 Optimal Design Solutions.** After having a predictive model, the next step in solving the optimization problem involves calculating the objective value  $u$  each time the design variables (i.e., the values of suction power and weight of Product 369) are explored. As illustrated in Algorithm 1, we first employ the trained ERGM model with the estimated model parameters  $\theta_{est}$  provided in Table 4 to simulate 100 networks  $Y_l$ . Within this set of networks, we examine each triadic closure  $y_{P_{369}, V_i, V_j}^m$  contained in the set  $S$ . Our objective is to count its occurrence across the 100 networks and compute its occurring ratio, representing its probability of existence, denoted as  $\Pr(y_{P_{369}, V_i, V_j}^m)$ . Given that  $\Pr(y_{P_{369}, V_i, V_j}^m)$  typically exhibits a skewed distribution with most values low, we set the median of these probabilities as the threshold to determine the existence of each triadic closure. This choice ensures a more accurate measure of central tendency and provides robustness to outliers [66]. Accordingly, the number of existing inter-brand triadic closures, equivalent to  $N_{R1}^{P_{369}}$ , is obtained and incorporated into Eq. (4) to compute the objective value. Next, as described in Algorithm 2, the evaluation of the objective function is built into the GA [56] which helps to find the optimal level of suction power and weight with the constraint on price to maximize the objective value. The initialization of the GA algorithm is detailed in Algorithm 2, where “*popSize*” denotes the population size in each round of search, “*maxiter*” indicates the defined maximum number of generations to run before the GA search stops, and “*run*” means the maximum number of consecutive generations for which the best

objective value (fitness) has no improvement, leading to the termination of the GA search [56].

**Algorithm 1 Objective value calculation**

```

1: Given  $V, S, x_s^{P_{369}}, x_w^{P_{369}}, x_p^{P_{369}}, \theta_{est}$ 
2: Initiate  $L = 100$ 
3: Simulate  $L$  networks  $Y_l, (l = 1, \dots, L)$  with the given  $x_s^{P_{369}}, x_w^{P_{369}}, x_p^{P_{369}}$ ,
   and estimated ERGM parameters  $\theta_{est}$ 
4: for  $m = 1$  to  $M$  do
5:   count = 0
6:   for  $l = 1$  to  $L$  do
7:     if  $y_{P_{369}, V_i, V_j}^m$  exists in  $Y_l$  then
8:       count = count + 1
9:     end if
10:  end for
11:   $\Pr(y_{P_{369}, V_i, V_j}^m) = \text{count}/L$ 
12: end for
13:  $\Pr_{\text{threshold}} = \text{Median}(\Pr(y_{P_{369}, V_i, V_j}^m))$ 
14: for  $m = 1$  to  $M$  do
15:    $N_{R1}^{P_{369}} = 0$ 
16:   if  $\Pr(y_{P_{369}, V_i, V_j}^m) > \Pr_{\text{threshold}}$  then
17:      $N_{R1}^{P_{369}} = N_{R1}^{P_{369}} + 1$ 
18:   end if
19: end for
20: Return  $u = \exp(0.316 + 0.117N_{R1}^{P_{369}} - 0.002(N_{R1}^{P_{369}})^2)$ 

```

**Algorithm 2 Optimization process**

```

1: Constraint  $x_p^{P_{369}}$ 
2: Variable  $X = [x_s^{P_{369}}, x_w^{P_{369}}]$ 
3: fitness = function( $X$ ) + objective value calculation( $X$ )
4: GA (type = “real-valued,”
5:   fitness,
6:   min = [ $x_{\text{suclow}}, x_{\text{weiglowl}}$ ],
7:   max = [ $x_{\text{suchigh}}, x_{\text{weighhigh}}$ ],
8:   popSize = 30, maxiter = 100, run = 15)
9: Summary (GA)

```

**3.5 Comparison Between the Traditional and Proposed Design Methods.** In this section, we compare and evaluate the design outcomes between the traditional and proposed design methods. The key difference between these methods is that the traditional approach optimizes product design by treating each product independently, relying solely on the relationship between the design objective (e.g., maximizing market share) and product attributes. In contrast, the proposed method also considers local dependencies between products, such as competition, during the optimization process. Two design cases were analyzed using both methodologies. Case 1 focuses on designing the suction power of Product 369 to enhance its market competitiveness, considering the constraints on weight and price. In case 2, we optimize both the

**Table 6 Case 2: negative binomial regression estimated result of  $u$  corresponding to  $X = [x_s, x_w]$** 

Independent variables	Est. Coeff.	Std. error
Intercept	0.764**	0.295
$x_s$	-0.317*	0.135
$x_p$	0.0004	0.0003
$x_w$	0.038	0.035
$x_w^2$	-0.004*	0.002
$x_s x_w$	0.028**	0.011

\*\*0.001 level of significance.

\*0.01 level of significance.

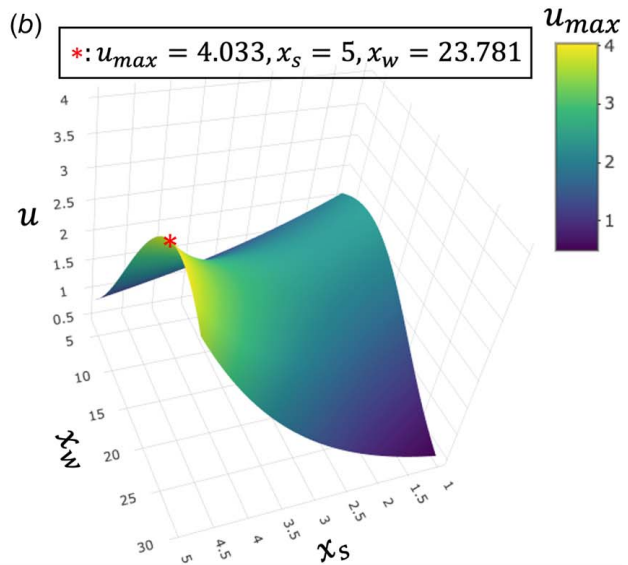
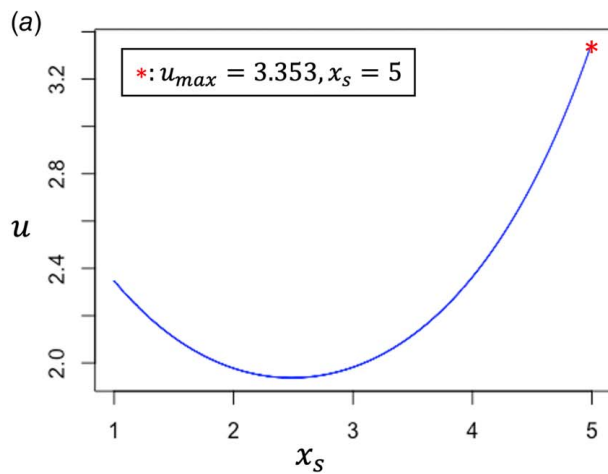


Fig. 7 Optimal design solutions obtained by the traditional design method

suction power and the weight of Product 369 with the objective of increasing its likelihood of being purchased, while keeping its price unchanged.

**Results of the Traditional Design Method:** Regarding the traditional method, we adhere to the procedure conducted in Sec. 3.2 to directly estimate the relationship between the number of times product purchases  $u$  and the original design vector  $\mathbf{X}$  using a negative binomial regression model, without local network representation of competition relations between products. Specifically, for case 1,  $\mathbf{X} = [x_s]$ , and for case 2,  $\mathbf{X} = [x_s, x_w]$ . Furthermore, we incorporate the price ( $x_s$ ) into the model due to its role as the constrained variable. The estimated results for both cases are presented separately in Tables 5 and 6.

Next, in case 1, we take the original price and weight of Product 369 ( $x_p = \$284.98$ ,  $x_w = 15.6$  LB) into the estimated regression model in Table 5, aiming to search the maximum  $u$  value within the specified suction power design range  $[1, 5]$ .<sup>4</sup> In case 2, we maintain the original price constraint while relaxing the weight constraint. Our goal is to identify the maximum  $u$  value within the design space defined by the suction power range  $[1, 5]$  and the weight range  $[3.34$  LB,  $29.3$  LB]. According to Fig. 7(a), the highest  $u = 3.353$  occurs at  $x_s = 5$ . In Fig. 7(b), the highest  $u = 4.033$  is achieved when  $x_s = 5$  and  $x_w = 23.781$  LB.

**Results of the Proposed Design Method:** Following the methodology outlined in the preceding subsections, we execute the algorithm detailed in Sec. 3.4 to explore optimal design solutions. Figure 8 shows the converged search process for optimal values. In case 1, the search terminates in the 15th generation as there is no improvement in the best objective value for 15 consecutive generations. Throughout these generations, we identify ten optimal suction power values at  $[2.676, 2.680, 2.684, 2.688, 2.692, 2.696, 2.704, 2.708, 2.712, 2.716]$  (decrease by  $1.284$ – $1.324$  from its original design), which corresponds to the best objective value of  $u = 6.493$ . In case 2, convergence is achieved by the 15th generation as well. We identify three sets of optimal solutions

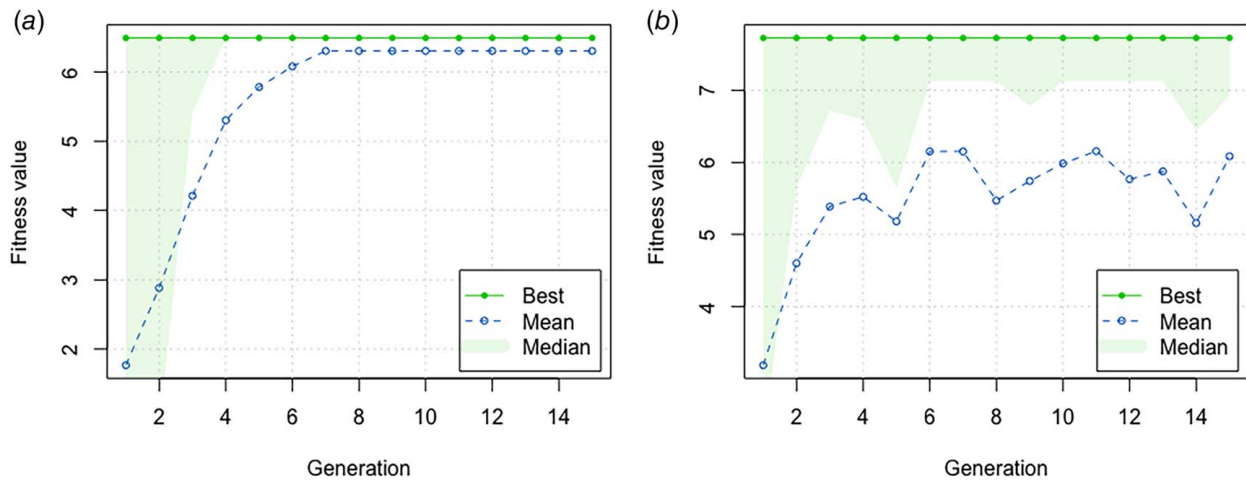
characterized by weight and suction power values  $[x_s, x_w]$ :  $[4.291, 22.317$  LB],  $[4.445, 28.457$  LB], and  $[1.985, 20.078$  LB]. These solutions align with the best objective value of  $7.729$ . Lastly, according to the mean value curves of the iterative search processes for both cases, we can observe that the search process in case 2 is more fluctuant. This fluctuation could be attributed to the fact that the design space of case 2, corresponding to two product attributes, is 2D and therefore much larger and more complex than the 1D design space of case 1.

**Results Comparison:** A comparison of the final results between the traditional and proposed design methods is summarized in Table 8 (the columns corresponding to the traditional design method and proposed design method for  $g(\mathbf{y}(\mathbf{X})) = [N_{R1}]$ ). According to the table, the proposed design method achieves objective values approximately twice as high as those of the traditional design method in both cases. For instance, in case 2, the proposed design method, considering local inter-brand triadic closure competition relationships, achieves an objective value of  $7.729$ . If translating to the number of times purchased, this is approximately eight times, about two times higher than the traditional design method ( $4.033$ ). Moreover, the objective values obtained using the proposed method for both cases significantly exceed those derived from applying the original design to the proposed algorithm. This further demonstrates the efficiency of the proposed design method. Lastly, the optimal design of case 2 using the proposed design method provides three design options, which offer varying trade-offs between suction power and weight, allowing for tailored solutions catering to different customer preferences while maximizing the product's market appeal.

**3.6 Extensibility of the Proposed Design Method.** In this section, we demonstrate the extensibility of the proposed design method by considering the second most important node role,  $R2$ , as shown in Table 2. Consequently, the derived local network-based design variable changes from  $g(\mathbf{y}(\mathbf{X})) = [N_{R1}]$  to  $g(\mathbf{y}(\mathbf{X})) = [N_{R1}, N_{R2}]$ . Accordingly, the optimal design problem is reformulated by estimating the relationship between  $u$  and the new  $g(\mathbf{y}(\mathbf{X}))$ .

**Optimal Design Reformulation and Solution:** The estimated result is provided in Table 7. The new local network-based design objective function for the derived variable is illustrated in Eq. (6). The number of times Product 369's involvement in node role  $R2$ ,  $N_{R2}^{P369}(\mathbf{y}(\mathbf{X}^{P369}))$ , can be expressed similarly to  $N_{R1}^{P369}(\mathbf{y}(\mathbf{X}^{P369}))$  in Eq. (5). In the network model shown in Fig. 5, 51 unique products, including Product 369, are from Dyson, resulting in 1225 potential

<sup>4</sup>In the dataset, vacuum cleaners of different brands or categories have different units for suction power. Two commonly used units are horsepower and airflow [67]. To solve the problem, we first unify the suction powers of the same unit in the range  $[1, 5]$  without units. For example, if the original airflow interval is  $[21.2$  CFM,  $160$  CFM] (CFM: cubic feet per minute), we evenly divide it into five subintervals. Products with airflow in the first subinterval  $[21.2$  CFM,  $48.96$  CFM] are assigned a level value of 1, and the same operation applies to levels 2, 3, 4, and 5. In cases with multiple suction power units for the same product, the final level value is the average of the level values converted from multiple units.



**Fig. 8** Iterative search processes using a genetic algorithm to optimize two design cases. The processes terminate after 15 generations, respectively, with a convergence criterion of no improvement in the best objective value (fitness) for 15 consecutive generations. In the plot, the lower boundary of the shaded area represents the median fitness value, while the upper boundary corresponds to the best (maximum) fitness value. This shaded area delineates the range within which the fitness values of the top 50% of the population fall in each generation. Consequently, it visualizes the spread and variability of the fitness values within the upper half of the population. (a) Case 1: suction power design case and (b) case 2: suction power and weight design case.

intra-brand triadic closure competitions for Product 369. Keeping the rest of the settings unchanged, the updated optimization problem is provided in Appendix (Fig. 10). Solving this updated optimization problem follows the same logic as introduced in Sec. 3.4, with minor revisions to Algorithm 1 while keeping Algorithm 2 unchanged. Since most steps are the same, we do not repeat them here and have included the updated Algorithm 1 in Appendix.

$$u(\mathbf{g}(\mathbf{y}(X^{P369}))) = \exp(0.227 + 0.097N_{R1}^{P369} - 0.002(N_{R1}^{P369})^2 + 0.204N_{R2}^{P369} - 0.014(N_{R2}^{P369})^2) \quad (6)$$

**Results of the Extensibility Test:** Figure 9 shows the converged search process for optimal values. In case 1, the search terminates in the 15th generation as there is no improvement in the best objective value for 15 consecutive generations. Throughout these generations, we identify 18 optimal suction power values, as listed in Table 8 corresponding to column  $\mathbf{g}(\mathbf{y}(X)) = [N_{R1}, N_{R2}]$ , which decrease by 1.248–1.336 from the original design and correspond to the best objective value of 11.555. In case 2, convergence is achieved by the 16th generation. We identify 21 sets of optimal solutions characterized by weight and suction power values  $[x_w, x_s]$  given in Table 8. These solutions align with the best objective value of 12.525. Comparing these results with the traditional design method and the proposed method before extension (i.e., not including node role  $R2$  into account), the extended method achieves the highest objective values for both cases. For example,

**Table 7** Negative binomial regression estimated result of  $u$  corresponding to  $\mathbf{g}(\mathbf{y}(X)) = [N_{R1}, N_{R2}]$

Independent variables	Est. Coeff.	Std. error
Intercept	0.227**	0.077
$N_{R1}$	0.097***	0.017
$N_{R1}^2$	-0.002**	0.0005
$N_{R2}$	0.204***	0.056
$N_{R2}^2$	-0.014*	0.006

\*\*\*0.000 level of significance.

\*\*0.001 level of significance.

\*0.01 level of significance.

in case 2, highlighted in bold in Table 8, the number of times product purchases reaches around 13 when considering both inter-brand and intra-brand competitions. This is approximately three times higher than the traditional design method and five units more than the design that only considers inter-brand competition. This highlights the importance of a comprehensive understanding of the market competition environment for effective product design.

## 4 Discussion

In this section, we first discuss the generalizability of the proposed network-based design framework, and then discuss the limitations of the current work and suggest future directions for improvement.

**4.1 Generalizability of the Proposed Method.** The generalizability of the proposed method is represented in two aspects: (1) *generalizability in handling complexity*. The proposed method is flexible and can handle different levels of complexity in optimization design problems. As illustrated in Sec. 3.5, it can optimize a varying number of product attributes. Additionally, Sec. 3.6 demonstrates that the derived local network-based design variable  $\mathbf{g}(\mathbf{y}(X))$  can be adjusted to include different local network structures. This adaptability allows the model to manage a wide range of complexities, making it applicable to various scenarios. (2) *Generalizability across cases*. Beyond vacuum cleaner product design, the proposed method can be directly applied to other product designs, such as vehicles and cellphones, by incorporating market competition information into the product design process. Additionally, the method can be generalized to guide the design of other networked systems, such as transportation systems and power grids. For example, in a shared mobility system, each docked bike station can be defined as a node in a trip network, with directed links representing trips between stations. Using a network motif mining tool, significant travel patterns can be identified to formulate the derived local network-based design variable for each station. This variable, determined by original station design parameters like dock numbers, can incorporate significant user travel patterns into station capacity design, improving system performance, such as user satisfaction scores.

**Table 8 Design results comparison: traditional and proposed methods**

Traditional design method			Proposed design method					
			$g(\mathbf{y}(\mathbf{X})) = [N_{R1}]$			$g(\mathbf{y}(\mathbf{X})) = [N_{R1}, N_{R2}]$		
			Case 1: $\mathbf{X} = [x_s]$					
	$x_s$	$u$		$x_s$	$u$		$x_s$	$u$
Original design	4	2.364 <sup>a</sup>	Original design	4	0.854 <sup>a</sup>	Original design	4	2.146 <sup>a</sup>
Optimal design	5	3.353	Optimal design	[2.676, 2.680, 2.684, 2.688, 2.692, 2.696, 2.704, 2.708, 2.720, 2.724, 2.728, 2.732, 2.736, 2.740, 2.744, 2.748, 2.752]	6.493	Optimal design	[2.664, 2.668, 2.672, 2.676, 2.684, 2.692, 2.696, 2.704, 2.716, 2.712, 2.716]	11.555
			Case 2: $\mathbf{X} = [x_s, x_w]$					
	$x_s$	$x_w$	$u$	$[x_s, x_w]$	$u$		$[x_s, x_w]$	$u$
Original design	4	15.6	2.802	Original design [4, 15.6 LB]	0.854 <sup>b</sup>	Original design	Original design [4, 15.6 LB]	2.146 <sup>b</sup>
Optimal design	5	23.78	4.033	Optimal design [4.291, 22.317 LB] [4.445, 28.457 LB] [1.985, 20.078 LB]	7.729	<b>Optimal design</b>	<b>[2.904, 18.723 LB] [3.216, 18.760 LB] [3.525, 18.797 LB] [3.164, 18.723 LB] [3.236, 18.760 LB] [3.280, 18.797 LB] [3.200, 18.723 LB] [3.296, 18.760 LB] [3.316, 18.797 LB] [3.200, 18.723 LB] [3.296, 18.760 LB] [3.316, 18.797 LB] [3.156, 18.760 LB] [3.384, 18.760 LB] [3.408, 18.797 LB] [3.176, 18.760 LB] [2.976, 18.797 LB] [3.164, 19.093 LB] [3.188, 18.760 LB] [3.228, 18.797 LB] [3.372, 19.204 LB]</b>	<b>12.525</b>

<sup>a</sup>The  $u$  value corresponding to the original design is calculated by inputting the original design values of Product 369 into the estimated negative binomial regression model for the traditional design method or into the developed algorithm for the proposed design method.

<sup>b</sup>Corresponding to a specific  $g(\mathbf{y}(\mathbf{X}))$ , cases 1 and 2 share the same algorithm, resulting in the same value of  $u$  for the original design of both cases.

**4.2 Limitation.** The first limitation of the current work is the inadequate predictive accuracy of the ERGM obtained in Step 4. One key reason for this could be attributed to the data insufficiency. As stated in our previous study [59], the data for US household vacuum cleaners, including 945 customer responses to 612 unique vacuum cleaner models, are quite heterogeneous, i.e., most customers' preferred vacuum cleaners are very different from others. This makes the co-consideration network have insufficient links to train an effective ERGM for prediction. Inspired by the existing study [42], where a GNN model was trained using a dataset aggregated from more than 40,000 vehicle survey responses to predict the co-consideration network for the vehicle market system with an  $F1$ -score of 0.65, we propose two potential solutions. First, we could collect more data. With more customer responses, we believe that the accuracy of the model will be improved. Second, we could use advanced deep learning models to replace ERGM since Step 4 of the proposed methodology only requires a network predictive model, and we plan to test more advanced deep learning models such as GNN as the surrogate model that is expected to further improve prediction accuracy.

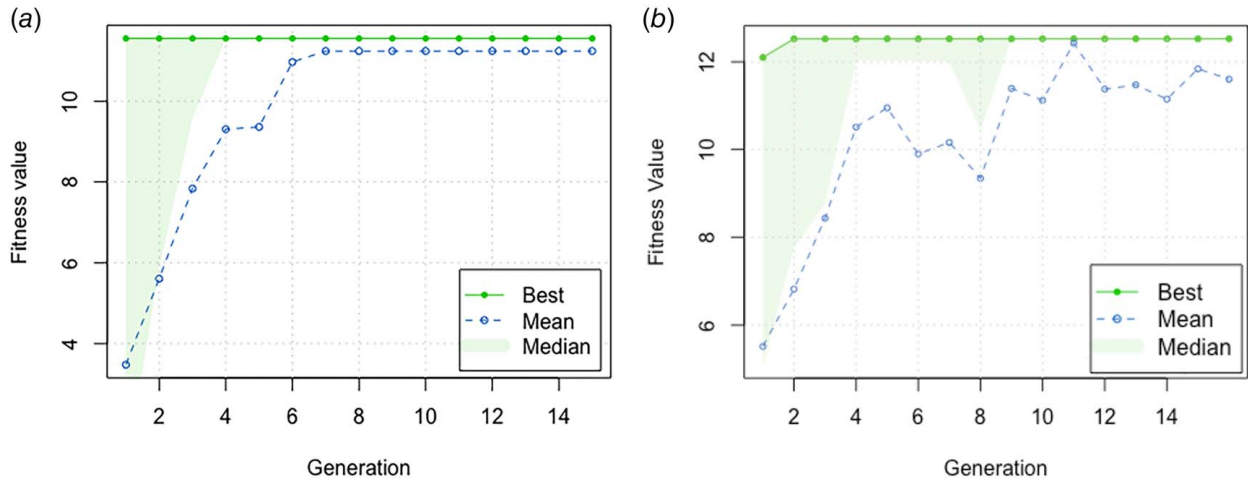
Another limitation of this study lies in the slow computational efficiency of Algorithm 1. The computer used in the experiments is equipped with an 11th-gen Intel Core CPU (i9-11900 2.50 GHz, 8 cores, 16 logical processors) and 32GB of RAM. Since ERGM is not compatible with graphics processing unit (GPU) calculation, we employ a parallel computing strategy utilizing 14 logical processors of the CPU. The computational time for each round of Algorithm 1 is approximately 4.3 min for only considering inter-brand competition and 4.8 min for considering both inter-brand and intra-brand competition. Consequently, each generation involving 30 populations of the genetic algorithm requires a total of 2.15 and 2.4 h, respectively. Therefore, solving the proposed optimal design problem and its extended version, which

encompasses 15 generations, requires 32.25 h and 36 h to complete the calculations. Moreover, the inefficiency of Algorithm 1 will also hinder the applicability of the proposed method to systems with large network sizes. To address this computational challenge, there are two potential directions to explore. One approach involves extending the current ERGM package [68] to make it compatible with GPU computing. Another direction is to utilize the aforementioned GNN model, which is GPU-compatible, as a replacement for ERGM in calculating the objective value.

## 5 Conclusion

In this study, we introduce a network-based system design framework, consisting of six key steps. The first step involves generating a network representation for the complex systems. In the second step, we perform significant local network mining and articulate the local network-based design goal, which is an essential step in integrating interdependencies between individual entities into the design process. Subsequently, in Step 3, we formulate an optimization problem based on the proposed local network-based design goal. Moving on to the fourth step, we develop a network predictive model as a surrogate to evaluate the system objective to prepare for solving the optimization problem. In Step 5, we integrate the predictive model into optimization algorithms, such as the genetic algorithm, to address the optimization problem outlined in Step 3. Here, the predictive model plays a key role in updating the objective value, while the genetic algorithm is employed to search for the optimal objective value. Finally, the obtained optimal design solution is utilized to recompute the objective value for validation.

To demonstrate the applicability of our approach in real-world scenarios, we present a case study on the US household vacuum cleaner market. The objective is to optimize a specific product model's design attributes to increase its sales. Following the



**Fig. 9** Iterative search processes using a genetic algorithm to optimize two design cases. The processes terminate after 15 and 16 generations, respectively, with a convergence criterion of no improvement in the best objective value (fitness) for 15 consecutive generations. (a) Case 1: suction power design case and (b) case 2: suction power and weight design case.

proposed method, we first model the vacuum cleaner market competition as a unidimensional co-consideration network. Next, we employ network motif theory to identify three significant local competition patterns and define the derived local network-based design variable based on the unique node positions in those identified competition motifs. This derived variable is a function of the original product design variables, including suction power, weight, and price (as a constrained variable). With the goal of maximizing the number of times a vacuum cleaner is purchased, we formulate a local network-based design objective function by estimating the relationships between purchase times and the derived network-based design variable using negative binomial regression. We then frame an optimization problem based on this objective function and solve it using a typical genetic algorithm procedure. In this process, the ERGM-based predictive model works as a surrogate model to evaluate the system objective whenever a new value of the design attributes (i.e., weight and suction power) is explored.

We demonstrate the efficiency of our proposed design method by comparing it with the traditional design method. The results show that the optimal values of suction power and weight found by the proposed method can significantly enhance the number of times product is purchased, achieving about twice the increase compared to the traditional design method. Additionally, we demonstrate the extensibility of the proposed method by modifying the derived design variable to include more competition relations. The highest objective value illustrates the success of this extension and highlights the importance of comprehensively understanding the market competition environment for optimal product design.

To address the aforementioned limitations, an immediate plan is to develop a new graph neural network-based predictive model. This model will be seamlessly integrated into the optimization algorithm, helping to predict the network structure and compute the objective value.

### Acknowledgment

The authors acknowledge collaborators Neelam Modi, Jonathan Haris Januar, Michael T. Cardone, and Gracia Cosenza for their assistance in data collection, data processing, and the inputs provided during research meetings. We also greatly acknowledge the funding support from NSF CMMI #2005661 and #2203080.

### Conflict of Interest

There are no conflicts of interest.

### Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

### Appendix: Details of Optimization Problem Formulation and Solving for Extension Case $g(\mathbf{y}(X)) = [N_{R1}, N_{R2}]$

Figure 10 presents the formulated optimization problem for the extension case, where  $g(\mathbf{y}(X)) = [N_{R1}, N_{R2}]$ . The primary difference from the pre-update case is in the objective function, which now reflects the estimated relationship between the number of times product is purchased  $u$  and the updated  $g(\mathbf{y}(X)) = [N_{R1}, N_{R2}]$ .

Algorithm 3 calculates the objective value for the extension case. The key differences from Algorithm 1 include: (1) in row 1, Algorithm 1, which contains all potential inter-brand triadic closures involving  $P369$ . In contrast, Algorithm 3 provides  $S_{R1}$  for all potential inter-brand triadic closures and  $S_{R2}$  for all potential intra-brand

### Optimization Problem Formulation

Maximize the number of times Product 369 purchases.

$$\max u(g(\mathbf{y}(X^{P369}))) = \max \exp(0.227 + 0.097N_{R1}^{P369}(\mathbf{y}(X^{P369})) - 0.002(N_{R1}^{P369}(\mathbf{y}(X^{P369})))^2 + 0.204N_{R2}^{P369}(\mathbf{y}(X^{P369})) - 0.014(N_{R2}^{P369}(\mathbf{y}(X^{P369})))^2)$$

$N_{R1}^{P369}(\mathbf{y}(X^{P369}))$ : the number of times Product 369 is involved in node role  $R1$ .

$N_{R2}^{P369}(\mathbf{y}(X^{P369}))$ : the number of times Product 369 is involved in node role  $R2$ .

$X^{P369} = [x_s^{P369}, x_w^{P369}]$

S.t.\*  $x_s^{P369} \in [x_{suc\_low}, x_{suc\_high}] = [1, 5]$

$x_w^{P369} \in [x_{weig\_low}, x_{weig\_high}] = [3.34, 29.3]$

$x_p^{P369} = \$284.98$

\*: Given that Product 369 belongs to upright vacuum cleaner, we establish the design space's lower and upper bounds by utilizing the minimum and maximum suction power and weights observed among upright vacuum cleaners available in the market.

**Fig. 10** The optimization problem formulation corresponding to  $g(\mathbf{y}(X)) = [N_{R1}, N_{R2}]$

triadic closures involving  $P369$ . (2) From row 13 to row 21 in Algorithm 3, the existence probability of each intra-brand triadic closure is calculated. Correspondingly, rows 29–34 count the number of existing intra-brand triadic closures. (3) Row 35 in Algorithm 3 returns the objective value calculated by the updated objective function.

### Algorithm 3 Objective value calculation

---

```

1: Given  $V, S_{R1}, S_{R2}, x_s^{P369}, x_w^{P369}, x_p^{P369}, \theta_{est}$ 
2: Initiate  $L = 100$ 
3: Simulate  $L$  networks  $Y_l, (l = 1, \dots, L)$  with the given  $x_s^{P369}, x_w^{P369}, x_p^{P369}$ ,
   and estimated ERGM parameters  $\theta_{est}$ 
4: for  $m = 1$  to  $M_{R1}$  then
5:   count = 0
6:   for  $l = 1$  to  $L$  do
7:     if  $y_{P369, V_i, V_j}^m$  exists in  $Y_l$  then
8:       count = count + 1
9:     end if
10:  end for
11:   $\Pr(y_{P369, V_i, V_j}^m) = \text{count}/L$ 
12: end for
13: for  $m = 1$  to  $M_{R2}$ 
14:   count = 0
15:   for  $l = 1$  to  $L$ 
16:     if  $y_{P369, V_i, V_j}^m$  exists in  $Y_l$  then
17:       count = count + 1
18:     end if
19:   end for
20:    $\Pr(y_{P369, V_i, V_j}^m) = \text{count}/L$ 
21: end for
22:  $\Pr_{\text{threshold}} = \text{Median}(\Pr(y_{P369, V_i, V_j}^m))$ 
23: for  $m = 1$  to  $M_{R1}$  do
24:    $N_{R1}^{P369} = 0$ 
25:   if  $\Pr(y_{P369, V_i, V_j}^m) > \Pr_{\text{threshold}}$ 
26:      $N_{R1}^{P369} = N_{R1}^{P369} + 1$ 
27:   end if
28: end for
29: for  $m = 1$  to  $M_{R2}$ 
30:    $N_{R2}^{P369} = 0$ 
31:   if  $\Pr(y_{P369, V_i, V_j}^m) > \Pr_{\text{threshold}}$  then
32:      $N_{R2}^{P369} = N_{R2}^{P369} + 1$ 
33:   end if
34: end for
35: Return  $u = \exp(0.227 + 0.097N_{R1}^{P369} - 0.002(N_{R1}^{P369})^2) +$ 
    $0.204N_{R2}^{P369} - 0.014(N_{R2}^{P369})^2$ 

```

---

## References

- [1] Sheard, S. A., and Mostashari, A., 2009, "Principles of Complex Systems for Systems Engineering," *Syst. Eng.*, **12**(4), pp. 295–311.
- [2] Hazelrigg, G. A., 1998, "A Framework for Decision-Based Engineering Design," *ASME J. Mech. Des.*, **120**(4), pp. 653–658.
- [3] Chen, W., Hoyle, C., and Wassenaar, H. J., 2013, *Decision-Based Design: Integrating Consumer Preferences Into Engineering Design*, Springer, London.
- [4] Taguchi, G., 1987, *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs*, UNIPUB/Kraus International Publications, New York.
- [5] Sha, Z., Kannan, K. N., and Panchal, J. H., 2015, "Behavioral Experimentation and Game Theory in Engineering Systems Design," *ASME J. Mech. Des.*, **137**(5), p. 051405.
- [6] Rubinstein, R. Y., and Kroese, D. P., 2016, *Simulation and the Monte Carlo Method*, John Wiley & Sons, New Jersey.
- [7] Torres, L., Blevins, A. S., Bassett, D., and Eliassi-Rad, T., 2021, "The Why, How, and When of Representations for Complex Systems," *SIAM Rev.*, **63**(3), pp. 435–485.
- [8] Chen, W., Heydari, B., Maier, A. M., and Panchal, J. H., 2018, "Network-Based Modeling and Analysis in Design," *Des. Sci.*, **4**, p. e16.
- [9] Walsh, H. S., Dong, A., and Turner, I. Y., 2019, "An Analysis of Modularity as a Design Rule Using Network Theory," *ASME J. Mech. Des.*, **141**(3), p. 031102.
- [10] Li, Y., Wang, Z., Zhang, L., Chu, X., and Xue, D., 2017, "Function Module Partition for Complex Products and Systems Based on Weighted and Directed Complex Networks," *ASME J. Mech. Des.*, **139**(2), p. 021101.
- [11] Song, B., McComb, C., and Ahmed, F., 2022, "Assessing Machine Learnability of Image and Graph Representations for Drone Performance Prediction," *Proc. Des. Soc.*, **2**, pp. 1777–1786.

- [12] Fuge, M., Tee, K., Agogino, A., and Maton, N., 2014, "Analysis of Collaborative Design Networks: A Case Study of Openideo," *ASME J. Comput. Inf. Sci. Eng.*, **14**(2), p. 021009.
- [13] Sha, Z., Chaudhari, A. M., and Panchal, J. H., 2019, "Modeling Participation Behaviors in Design Crowdsourcing Using a Bipartite Network-Based Approach," *ASME J. Comput. Inf. Sci. Eng.*, **19**(3), p. 031010.
- [14] Wu, D., Rosen, D. W., Panchal, J. H., and Schaefer, D., 2016, "Understanding Communication and Collaboration in Social Product Development Through Social Network Analysis," *ASME J. Comput. Inf. Sci. Eng.*, **16**(1), p. 011001.
- [15] Li, N., Kramer, J., Gordon, P., and Agogino, A., 2018, "Co-author Network Analysis of Human-Centered Design for Development," *Des. Sci.*, **4**, p. e10.
- [16] Chaudhari, A. M., Gralla, E. L., Szajnfarber, Z., and Panchal, J. H., 2022, "Co-evolution of Communication and System Performance in Engineering Systems Design: A Stochastic Network-Behavior Dynamics Model," *ASME J. Mech. Des.*, **144**(4), p. 041706.
- [17] Piccolo, S. A., Lehmann, S., and Maier, A., 2018, "Design Process Robustness: A Bipartite Network Analysis Reveals the Central Importance of People," *Des. Sci.*, **4**, p. e1.
- [18] Ball, Z., and Lewis, K., 2018, "Observing Network Characteristics in Mass Collaboration Design Projects," *Des. Sci.*, **4**, p. e4.
- [19] Mane, M., DeLaurentis, D., and Frazho, A., 2011, "A Markov Perspective on Development Interdependencies in Networks of Systems," *ASME J. Mech. Des.*, **133**(10), p. 101009.
- [20] Hassannezhad, M., Cantamessa, M., Montagna, F., and Clarkson, P. J., 2019, "Managing Sociotechnical Complexity in Engineering Design Projects," *ASME J. Mech. Des.*, **141**(8), p. 081101.
- [21] Dave, T., and Layton, A., 2020, "Designing Ecologically-Inspired Robustness Into a Water Distribution Network," *J. Cleaner Prod.*, **254**, p. 120057.
- [22] Chatterjee, A., Helbig, C., Malak, R., and Layton, A., 2023, "A Comparison of Graph-Theoretic Approaches for Resilient System of Systems Design," *J. Comput. Inf. Sci. Eng.*, **23**(3), p. 030906.
- [23] Layton, A., Bras, B., and Weissburg, M., 2016, "Designing Industrial Networks Using Ecological Food Web Metrics," *Environ. Sci. Technol.*, **50**(20), pp. 11243–11252.
- [24] Panyam, V., Huang, H., Davis, K., and Layton, A., 2019, "Bio-inspired Design for Robust Power Grid Networks," *Appl. Energy*, **251**, p. 113349.
- [25] Watson, B. C., Malone, S., Weissburg, M., and Bras, B., 2020, "Adding a Detrital Actor to Increase System of System Resilience: A Case Study Test of a Biologically Inspired Design Heuristic to Guide Sociotechnical Network Evolution," *ASME J. Mech. Des.*, **142**(12), p. 121705.
- [26] Song, B., Srinivasan, V., and Luo, J., 2017, "Patent Stimuli Search and Its Influence on Ideation Outcomes," *Des. Sci.*, **3**, p. e25.
- [27] Ahmed, F., and Fuge, M., 2018, "Creative Exploration Using Topic-Based Bisociative Networks," *Des. Sci.*, **4**, p. e12.
- [28] 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**, p. e11.
- [29] Wang, M., and Chen, W., 2015, "A Data-Driven Network Analysis Approach to Predicting Customer Choice Sets for Choice Modeling in Engineering Design," *ASME J. Mech. Des.*, **137**(7), p. 071410.
- [30] Bi, Y., Xie, J., Sha, Z., Wang, M., Fu, Y., and Chen, W., 2018, "Modeling Spatiotemporal Heterogeneity of Customer Preferences in Engineering Design," *ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Quebec City, Quebec, Canada, Aug. 26–29.
- [31] 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**(3), pp. 668–677.
- [32] Zadbood, A., and Hoffenson, S., 2022, "Social Network Word-of-Mouth Integrated Into Agent-Based Design for Market Systems Modeling," *ASME J. Mech. Des.*, **144**(7), p. 071701.
- [33] Giffin, M., de Weck, O., Bounova, G., Keller, R., Eckert, C., and Clarkson, P. J., 2009, "Change Propagation Analysis in Complex Technical Systems," *ASME J. Mech. Des.*, **131**(8), p. 081001.
- [34] Xiao, Y., and Sha, Z., 2022, "Robust Design of Complex Socio-technical Systems Against Seasonal Effects: A Network Motif-Based Approach," *Des. Sci.*, **8**, p. e2.
- [35] Dey, A. K., Gel, Y. R., and Poor, H. V., 2019, "What Network Motifs Tell Us About Resilience and Reliability of Complex Networks," *Proc. Natl. Acad. Sci. USA*, **116**(39), pp. 19368–19373.
- [36] Park, K., and Kremer, G. E. O., 2019, "Topological Characterization of an Evolving Product Structure Network: A Case Study of Generational Smartphone Products," *ASME J. Mech. Des.*, **141**(9), p. 091103.
- [37] Giffin, M., De Weck, O., Bounova, G., Keller, R., Eckert, C., and Clarkson, P. J., 2009, "Change Propagation Analysis in Complex Technical Systems," *ASME J. Mech. Des.*, **131**(8), p. 081001.
- [38] Sha, Z., Cui, Y., Xiao, Y., Stathopoulos, A., Contractor, N., Fu, Y., and Chen, W., 2023, "A Network-Based Discrete Choice Model for Decision-Based Design," *Des. Sci.*, **9**, p. e7.
- [39] Sha, Z., Huang, Y., Fu, J. S., Wang, M., Fu, Y., Contractor, N., and Chen, W., 2018, "A Network-Based Approach to Modeling and Predicting Product Coconsideration Relations," *Complexity*, **2018**(1), pp. 1–14.
- [40] Xiao, Y., Ahmed, F., and Sha, Z., 2023, "Graph Neural Network-Based Design Decision Support for Shared Mobility Systems," *ASME J. Mech. Des.*, **145**(9), p. 091703.
- [41] Wu, J., and Wang, P., 2023, "Generative Design for Resilience of Interdependent Network Systems," *ASME J. Mech. Des.*, **145**(3), p. 031705.

- [42] Ahmed, F., Cui, Y., Fu, Y., and Chen, W., 2022, "Product Competition Prediction in Engineering Design Using Graph Neural Networks," *ASME Open J. Eng.*, **1**, p. 011020.
- [43] Ferrero, V., DuPont, B., Hassani, K., and Grandi, D., 2022, "Classifying Component Function in Product Assemblies With Graph Neural Networks," *ASME J. Mech. Des.*, **144**(2), p. 021406.
- [44] Li, K., Gao, Y., Zheng, H., and Tan, J., 2021, "A Data-Driven Methodology to Improve Tolerance Allocation Using Product Usage Data," *ASME J. Mech. Des.*, **143**(7), p. 071101.
- [45] Xiao, Y., Cui, Y., Raut, N., Januar, J., Koskinen, J., Contractor, N., Chen, W., and Sha, Z., 2022, "Information Retrieval and Survey Design for Two-Stage Customer Preference Modeling," *Proc. Des. Soc.*, **2**, pp. 811–820.
- [46] Arora, J. S., 2004, *Introduction to Optimum Design*, Elsevier, Amsterdam.
- [47] Martins, J. R., and Ning, A., 2021, *Engineering Design Optimization*, Cambridge University Press, Cambridge.
- [48] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U., 2002, "Network Motifs: Simple Building Blocks of Complex Networks," *Science*, **298**(5594), pp. 824–827.
- [49] Schwöbbermeyer, H., 2008, *Network Motifs*, chapter 5, John Wiley & Sons Ltd, Hoboken, NJ, pp. 85–111.
- [50] Robins, G., Pattison, P., Kalish, Y., and Lusher, D., 2007, "An Introduction to Exponential Random Graph ( $p^*$ ) Models for Social Networks," *Soc. Netw.*, **29**(2), pp. 173–191.
- [51] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G., 2008, "The Graph Neural Network Model," *IEEE Trans. Neural Netw.*, **20**(1), pp. 61–80.
- [52] Butts, C. T., Morris, M., Krivitsky, P. N., Almquist, Z., Handcock, M. S., Hunter, D. R., Goodreau, S. M., and de Moll, S. B., 2014, *Introduction to Exponential-Family Random Graph (ERG Or  $p^*$ ) Modeling With ERGM*, European University Institute, Florence. <http://cran.r-project.org/web/packages/ergm/vignettes/ergm.pdf>
- [53] Hunter, D. R., 2007, "Curved Exponential Family Models for Social Networks," *Soc. Netw.*, **29**(2), pp. 216–230.
- [54] Morris, M., Handcock, M. S., and Hunter, D. R., 2008, "Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects," *J. Stat. Softw.*, **24**(4), p. 1548.
- [55] Goodfellow, I., Bengio, Y., and Courville, A., 2016, *Deep Learning*, MIT Press, Cambridge.
- [56] Whitley, D., 1994, "A Genetic Algorithm Tutorial," *Stat. Comput.*, **4**, pp. 65–85.
- [57] Kennedy, J., and Eberhart, R., 1995, "Particle Swarm Optimization," *IEEE Proceedings of ICNN'95 - International Conference on Neural Networks*, Vol. 4, Perth, WA, Australia, Nov. 27– Dec. 1, IEEE, pp. 1942–1948.
- [58] Xiao, Y., Cui, Y., Raut, N., Januar, J. H., Koskinen, J., Contractor, N., Chen, W., and Sha, Z., 2023, "Survey Data on Customer Two-Stage Decision-Making Process in Household Vacuum Cleaner Market," *Texas Data Repository*.
- [59] Xiao, Y., Cui, Y., Cardone, M. T., Chen, W., and Sha, Z., 2023, "Product Competition Analysis for Engineering Design: A Network Mining Approach," *Conference on Systems Engineering Research. CSER 2023*, Hoboken, NJ, Mar. 16–17.
- [60] Wernicke, S., and Rasche, F., 2006, "Fanmod: A Tool for Fast Network Motif Detection," *Bioinformatics*, **22**(9), pp. 1152–1153.
- [61] Elhai, J. D., Calhoun, P. S., and Ford, J. D., 2008, "Statistical Procedures for Analyzing Mental Health Services Data," *Psychiatry Res.*, **160**(2), pp. 129–136.
- [62] Chai, T., and Draxler, R. R., 2014, "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)?—Arguments Against Avoiding RMSE in the Literature," *Geosci. Model Dev.*, **7**(3), pp. 1247–1250.
- [63] Bozdogan, H., 1987, "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions," *Psychometrika*, **52**(3), pp. 345–370.
- [64] Kotler, P., Keller, K. L., Ancarani, F., and Costabile, M., 2014, *Marketing Management 14/e*, Pearson, Hoboken, NJ.
- [65] Dutka, A. F., and Hansen, H. H., 1991, *Fundamentals of Data Normalization*, Addison-Wesley Longman Publishing Co., Inc, Boston, MA.
- [66] Von Hippel, P. T., 2005, "Mean, Median, and Skew: Correcting a Textbook Rule," *J. Stat. Educ.*, **13**(2), pp. 965–971.
- [67] Vacuum Cleaner Specifications, <https://www.bestvacuum.com/pages/vacuum-cleaner-specifications>, Accessed January 1, 2024.
- [68] Krivitsky, P. N., Hunter, D. R., Morris, M., and Klumb, C., 2023, "ERGM 4: New Features for Analyzing Exponential-Family Random Graph Models," *J. Stat. Softw.*, **105**(6), pp. 1–44.