Optimization of dynamic product offerings on online marketplaces: A network theory perspective

Zhou, Meihua; Angelopoulos, Spyros; Ou, Carol; Liu, Hongwei; Liang, Zhouyang

Published in:
Journal of Management Information Systems

Publication date:
2023

Document Version
Early version, also known as pre-print

Citation for published version (APA):
Zhou, M., Angelopoulos, S., Ou, C., Liu, H., & Liang, Z. (in press). Optimization of dynamic product offerings on online marketplaces: A network theory perspective. *Journal of Management Information Systems*.

OPTIMIZATION OF DYNAMIC PRODUCT OFFERINGS ON ONLINE MARKETPLACES: A NETWORK THEORY PERSPECTIVE

Meihua Zuo, Huizhou University, China

Spyros Angelopoulos, Durham University, United Kingdom

Carol X.J. Ou, Tilburg University, The Netherlands

Hongwei Liu, Guangdong University of Technology, China

Zhouyang Liang, Guangdong University of Technology, China

ABSTRACT

The fierce competition amongst brands on online marketplaces makes the optimization of offerings within this context a significant challenge. To address this challenge, we draw upon network theory and model the degree of competition through consumers' consideration sets. We use a large empirical dataset from one of the biggest online marketplaces to explore the dynamic relationship between network position and the degree of competition, and we depict the redistribution of market share of related offerings after adjusting their array. In doing so, we provide a theoretical reference on *when* and *how* brands should optimize their product offerings on online marketplaces. We further demonstrate that intra-brand cannibalization relations have a significantly greater impact on the degree of competition compared to inter-brand ones, while intra-brand cannibalization relations represent the main reason for fluctuations in the degree of competition. Hence, contrary to existing theoretical insights and practical intuitions, our findings demonstrate that brands should minimize the number and heterogeneity of their offerings within a market segment to increase their sales on online marketplaces.

Keywords: *Online marketplaces, dynamic competition analysis, clickstream data, spatial auto-regressive model, network analysis*

INTRODUCTION

The recent pandemic has drastically changed contemporary social and economic life, while the related implemented policies around the globe have put enormous pressure on the e-commerce resilience [62]. The resulting uncertainties have, thus, increasingly led brands to migrate their business endeavors to online marketplaces; for instance, the largest retailer in the United States has currently more than 150,000 brands in its online marketplace, almost four times more than before the pandemic[1].

Such an aggregation of brands has naturally led to fiercer competition on online marketplaces. To improve their degree of competition on online marketplaces, therefore, brands broadly adopt the competitive strategy of product proliferation [52, 57], which can attract consumers from rivals and increase revenue as well as market share, leading to a further increase in the degree of competition amongst brands (inter-brand competition) [52]. Product proliferation, however, can also result in a cannibalization effect, where products of a brand can be perceived as substitutes by consumers (intra-brand cannibalization) [42]. A stronger cannibalization effect can lead to a long tail in the sales distribution of a brand [15], increasing production and inventory costs, while also affecting the degree of inter-brand competition. Consequently, optimizing product offerings to enhance the degree of competition on online marketplaces is becoming increasingly vital for brands. Concurrently, rapid technological progress enables brands to shorten the cycle of product renewal. For example, Zara—a fast fashion brand—renews its product line twice a week[2]. Therefore, if the decision support aids of online marketplaces continue to follow traditional approaches for analyzing the degree of competition of a brand [54], they might misadjust production plans, which can lead to increased operational risks, and a reduction in sales [21]. To address the tensions of that phenomenon, we specifically explore:

How can brands dynamically balance intra-brand cannibalization and inter-brand competition in product offerings to enhance their degree of competition on online marketplaces?

To answer the research question of our study and ensure the generalizability, and applicability of our findings, we draw upon network theory, and we use product consideration sets of consumers from multiple product categories. The literature to date has mostly used supply-side information to assess the degree of competition between brands, such as the market share [39] or performance [97]. For instance, Li et al. [52] used data from the industry of personal computers and adopted an econometric approach to balance the effects of intra-brand cannibalization and inter-brand competition. Such approaches, however, are not consistent with the contemporary needs of brands on online marketplaces, as the time and economic cost of obtaining information on the supply side are relatively high, and such information is often outdated, not reflecting the real-time degree of competition of brands. Recent work has used the product consideration sets of consumers to identify competitive structures for

[1]https://www.statista.com/topics/4827/online-marketplaces
[2] https://www.henryharvin.com/blog/5-hidden-facts-about-zara-the-six-sigma-process

brands and products from a market demand perspective. For example, Ringel and Skiera [75] used product consideration sets of consumers to explore asymmetric competitive networks and identify distinct submarkets. An asymmetric competition network is defined according to the direction of competition between brands [1, 27, 58]; when, for instance, the competition between brands B_1 and B_2 is not equal to the competition between brands B_2 and B_1, then they have an asymmetric competitive relationship [75]. We extend the line of research on product consideration sets of consumers through an asymmetric competitive network perspective, by focusing on the dynamic change of the market segment to which a brand or a product belongs, and specifically addressing how the network position of brands can dynamically affect their degree of competition.

The extant information systems (IS) literature to date has used product consideration sets of consumers to i) model their time-variant clicking behavior from a single- [41] or multiple-website perspective [60], ii) demonstrate the relation between the browsing behavior and purchase intention of consumers [60, 78, 96], and iii) analyze the market structure and competition [27, 32]. This line of research treats product consideration sets of consumers as *static* [30, 52, 75]. The *temporal* dimension of consumers' behavior, however, can reflect changes in their shopping process and provide a more consistent perspective on brand competition, thereby providing novel insights for optimizing product offerings, enhancing the degree of competition, and unearthing factors that influence them over time. We use the product consideration sets of consumers in different periods to measure the temporal dimension of their behavior on online marketplaces. In doing so, we extend the relevant discussions on the IS literature [30, 52, 75] by using consideration sets of consumers over time to dynamically balance *intra-brand cannibalization* and *inter-brand competition*, as well as to show the dynamic change of the market segment to which the brand and product belongs to.

To explore how the market position of brands can dynamically affect their degree of competition on online marketplaces, we draw upon network theory; specifically, we use the network measures of *structural holes* and *centrality indexes*. According to network theory, a structural hole is a non-redundant connection between two nodes [17, 18], and describes a specific node that has direct connections with some nodes, but no connection or discontinuity with others. A brand in a structural hole refers to its position between other brands on an online marketplace, indicating that consumers rarely search for it at the same time along with others in a focal market segment, but they do search for it at the same time along with others in other segments. Consequently, although the focal brand is less threatened by competition in a specific market segment, it can be threatened in others. Prior studies on structural holes have shown that nodes can gain advantages through brokerage opportunities created by such a lack of ties among nodes [77, 88]. Other studies, however, show that nodes can benefit from a confidential relationship with others, and the degree of such confidentiality is often measured by centrality indexes. The most commonly used centrality indexes are degree centrality, and betweenness centrality [34, 35, 92]. Degree centrality measures the control range of nodes; nodes with a higher degree centrality have a stronger influence. Since degree centrality represents the direct connection with a specific node, it reflects the impact of nodes on the local market [69], which refers to the market of a brand and its direct competitors in which a consumer navigates from one to the other. Betweenness centrality represents the positioning of a node as an intermediary and measures the number of shortest paths passing through a focal node, without necessarily having direct ties [35]. Betweenness centrality, thus, measures the impact of nodes on the global market, which refers to the network of direct and indirect competitors. High centrality indexes show that consumers often search for a brand along with other brands, indicating its popularity in the market; degree centrality reflects the local market and betweenness centrality reflects the global one.

We use the span of structural holes, betweenness centrality, and degree centrality as independent variables (IV) on a dataset of 6,549,484 records over a period of 10 weeks from one of the biggest online marketplaces in Asia. We explore the degree of brand competitive dynamics through the time-variant clicking behavior of consumers, using clickstream data [43, 87], which can objectively reflect the underlying interests and preferences of consumers on online marketplaces [41, 60]. To address the issue of interdependence in the network [73], we use a spatial auto-regressive (SAR) model [91], and we develop a unified framework to assess intra- and inter-brand competition. We further focus on how the position of a brand in a competitive network can dynamically affect its degree of competition by modeling the degree of inter-brand competition as a function of network attributes. Contrary to established theoretical insights as well as current practical intuitions and market practices [85], we demonstrate that brands should *minimize* heterogeneity in product offerings on online marketplaces to increase sales. We find that intra-brand product relations have a greater impact on sales, and they are the main reason for competition fluctuations. Consequently, our findings bear novel insights into dynamic brand competition and the positioning of brands on online marketplaces.

The rest of our paper is organized as follows. In section 2, we first present the theoretical background of our study on asymmetric competition, implicit consumers' preferences, consumers' clicking behavior, networks, and the relationships amongst them. Second, in this section, we then present the theoretical background of our study on networks, as well as the relationships between networks and asymmetric competition. In section 3, we proceed to present the construction of the dynamic networks of our study, where we use consideration sets of consumers to quantify intra- and inter-brand competition. In section 4, we discuss the reduction of the intra-brand cannibalization effect. In section 5, we demonstrate how the network position of brands can dynamically affect their degree of competition in the context of online marketplaces. In section 5, we discuss the findings and contributions of our study, while in section 6 we conclude the paper with the implications of our study for both theory as well as practice and we delineate an agenda for future research on the topic.

THEORETICAL BACKGROUND

Clicking Behavior and Implicit Preferences

Online marketplaces enable consumers to simultaneously compare the offerings of multiple brands, leading to more intense competition than in offline settings [98]. In this context, the *shopping process* can be defined as the period between the initial searches for a product and the final purchase [44]. Consequently, brands within the same product category that co-occur within a *shopping process* can be perceived as competitors [24, 46, 65, 67, 75, 98]. Such a shopping process leaves behind trails in the form of clickstream data, which can be an effective resource for studying brand competition along with the interests and implicit preferences of consumers [12, 16]. Clickstream data represent electronic records of consumers' objective activities [16] and reflect their online browsing behavior without the need for introspection. Therefore, consumers' online browsing behavior can indirectly reflect their implicit preferences. The implicit preferences of consumers refer to those they do not need to introspect [56]; that is, consumers are not aware of their implicit preferences, or cannot accurately express them. The implicit preference of consumers can reveal more accurately their interests [33, 56], which can enhance the degree of competition for brands [20, 21].

The literature has broadly explored the heterogeneity and degree of competition between pairs of brands [89], with only a few studies focused on optimizing the product offerings of brands in competitive environments [52], such as online marketplaces. This line of research shows that: i) it is necessary to optimize product offerings in competitive environments [31] such as online marketplaces, ii) there is a consistent inverse U-shaped relationship between the degree of competition and the length of the product line of a brand [38], and iii) the degree of brand competition differs based on the category of products [75]. One of the key challenges in analyzing brand competition on online marketplaces is the increasing number of competing products since traditional methods such as scanner panels are not practically feasible for product categories containing thousands of products [67]. Prior studies have measured the implicit preferences of consumers by surveying the average number of their searches for each product before a purchase [52]. Such data, however, can be limited due to the cognitive abilities of consumers, as it can be difficult to recall what they have searched for in the past during their online shopping sessions. Especially when the number of products searched for is large, the accuracy of surveys gradually decreases [75]. In addition, the collection of survey data can be costly as well as time-consuming and cannot reflect the *real-time* preferences of consumers.

Data about the online clicking behavior of consumers can reflect their implicit preferences and can be used to construct their consideration sets before completing a purchase [65]. Since consideration sets are the arbiters of competitive relationships [72], they can enable the discovery of competitive market structures [70]. Clickstream data can capture the instantaneous actual behavior of consumers, which is more reliable than survey and panel scan data [68]. Consequently, the implicit preference of consumers that are captured by clickstream data is the most direct, effective, and objective reflection of potential asymmetric competitive relationships amongst brands. Prior studies have used online searching, clicking, browsing, and reviewing behaviors to reflect the implicit preferences of consumers, and to represent the potential competition relations of brands [28, 46, 51, 52, 67]. Amongst them, Ringel and Skiera [75] used data of consumer search from product- and price-comparison websites to analyze asymmetric intra- and inter-brand competition based on such consideration sets and to identify different submarkets. Less explored within this line of research, however, remain the *dynamic* identification of asymmetric intra- and inter-brand competition, as well as the in-depth analysis of their influencing factors for product positioning on online marketplaces. In further clarifying this lacuna, we present the similarities and differences of our study with the relevant prior ones in the literature on intra- and inter-brand competition networks, in Appendix I. We, therefore, approach intra- and inter-brand competition networks in line with Ringel and Skiera [75], while we further contribute to this line of research through the consideration of the *dynamic* asymmetric competition in various time-windows, and by further capturing the network-related factors that can influence the degree of asymmetric competition on online marketplaces. Such contributions are timely and methodologically vital for IS research, especially in the context of online marketplaces, as novel insights on these long-standing issues could lead to the reduction of production and inventory costs for brands, which is conducive to enhancing their degree of competition.

Asymmetric Competition Networks

We draw upon the foundations of network theory to explore the competing relations among brands, and to capture the factors influencing their degree of asymmetric competition. Network theory provides a robust way to measure the positional advantages of nodes within a network [73], and has been widely adopted for studying organizations [12, 25, 47], their strategy [37], their innovation processes [36, 40], as well as their operations management [19]. In the context of our study, for the inter-brand network, brands represent the nodes of the network, while the ties indicate that the corresponding brands have been considered by consumers within a shopping process. The weight of the ties indicates the probability of brands being jointly considered by consumers. For the intra-brand networks, the products represent the nodes, while the ties indicate that the corresponding product has been considered by consumers within a shopping process. The weight of the ties indicates the probability of products being jointly considered by consumers.

For the inter-brand competition network, a brand in a structural hole refers to its bridging position between distinct groups of brands on an online marketplace, which indicates that consumers rarely consider a focal brand along with others in this market segment at the same time, but they consider this brand along with others in other market segments. This means that although the focal brand is less threatened by competition in the market segment, it is threatened by competition in other market segments. Thus, opposing to the bridging role of structural holes in organizations for social relationships [17, 18], their role in the context of our study is expected to be negative. Consequently, brands residing in many structural holes are rarely considered with others, attracting fewer consumers, and as a result, achieving a lower volume of sales. Centrality is one of the most established network measures [34, 35], as it reveals the structural advantages of the position of a node [59]. For the inter-brand competition network, high centrality indicates that consumers consider the specific brand along with others with a high frequency, which means that the specific brand is well-known in the market segment. Thus, the combined use of structural holes and centrality measures can enable us to identify *which brands are more competitive* and to study the impact of their position in the inter-brand competition network on their degree of competition over time. *We, thus, argue that brands residing close to each other in such networks have stronger competing relations, due to a higher number of consumers that concurrently compare the utility of their offerings during the shopping process*.

To explore local and global market influence, we use betweenness and degree centrality [69]. Betweenness centrality measures the number of shortest paths passing through a focal node, which do not necessarily have direct ties connected to it [34, 35]. In the inter-brand network, such a focal node indicates that the brand is relatively close to the target brand, that is, consumers need to refer to attributes of the focal brand when searching for the target one. Although consumers do not make comparisons with target brands in a certain market segment frequently, they do refer to them in the global market. Betweenness, therefore, delineates the centrality of the entire market of related offerings. *We argue, therefore, that a brand with high betweenness centrality is more closely linked to the target brand, suggesting a higher exposure of the brand and, therefore, higher sales volume.*

Concurrently, degree centrality represents the number of nodes directly connected to a focal node [69]. In contrast to betweenness centrality, therefore, the degree of a node describes its local centrality within a certain market segment. In the context of our study, therefore, *we argue that the greater the value of a focal brand, the more traffic it generates, acquiring, thus, higher sales volume.*

DYNAMIC NETWORK CONSTRUCTIONS

To construct the dynamic networks of our study, we first identify the intra- and inter-brand competition relations based on the clickstream data, and then we model the degree of dynamic competition for each brand, as visually depicted in Figure 1. In doing so, our approach consists of four steps, as visually depicted in Figure 2: i) obtain data on consumers' sequential clicking and purchasing behavior for a certain category of products within a fixed period; ii) aggregate the data for each consumer in a shopping process to form a consideration set; iii) construct the asymmetric competition matrix by combining all the consideration sets of consumers; the matrix represents the probability that a particular brand is considered by consumers; and iv) model the asymmetric intra-brand relationships at the product level as well as the asymmetric inter-brand competition relationships at the brand level.

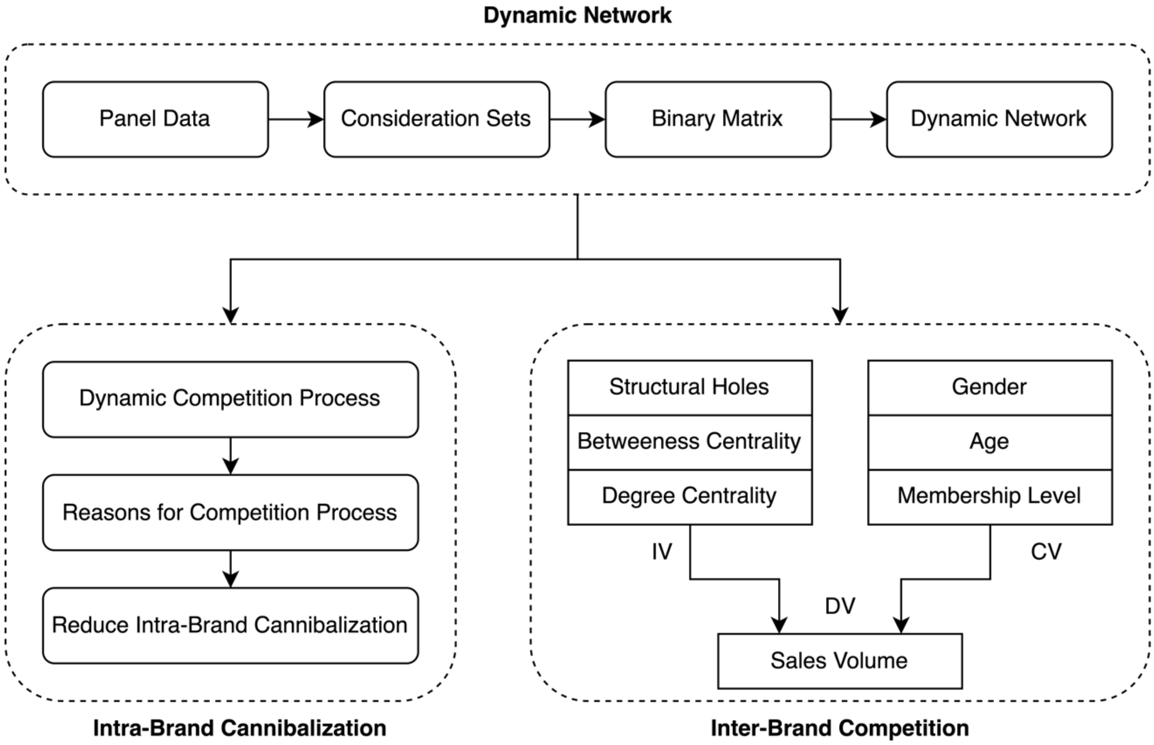

Figure 1: *Research Process*

Note: IV: Independent Variables; CV: Control Variables; DV: Dependent Variable.

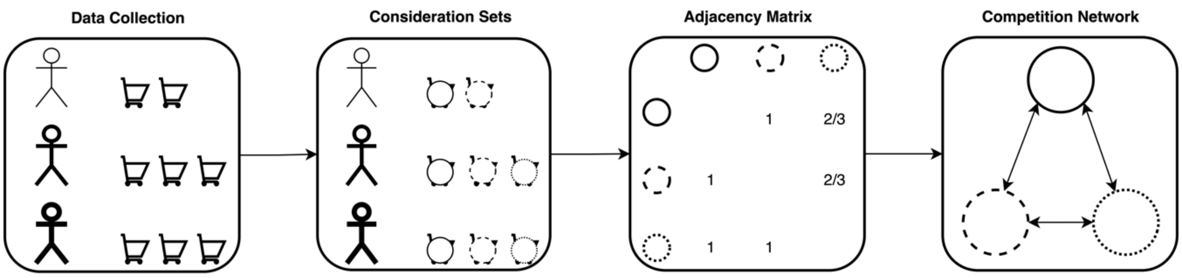

Figure 2: *Identification Steps for Asymmetric Relationships*

Data Collection

We obtained data on the online clicking and purchasing behavior of consumers for a period of 10 weeks from one of the biggest integrated online retail marketplaces in Asia, which has more than 300 million consumers and its core categories are electrical appliances, computers, and mobile phones. The longitudinal dataset contained: i) consumer sequential clicking behavior in the same product category;

ii) the log value of each brand's purchasing data; iii) consumer demographics (see Table 1), and iv) the attributes of clicked or purchased product (i.e., product name, brand name, and product category). With privacy in mind, the dataset was pseudonymized with the use of a coding system.

Table 1: *Descriptive Statistics of Demographic and Products*

Characteristic			Number
Demographic	**Gender**	**Male**	50,071
		Female	4,198
	Age	**Under 25 years old**	5,985
		26-35	26,530
		36-45	16,640
		46 years old or older	5,114
	Membership level	**Level 1**	4,548
		Level 2	10,647
		Level 3	16,701
		Level 4	22,373
Product	**Products included in each brand**	**Max**	198
		Min	1
		Mean	31.26
	Purchases by consumers	**Max**	17
		Min	1
		Mean	1.13
	Products purchased by consumers	**Max**	15
		Min	1
		Mean	1.08
	Brands purchased by consumers	**Max**	8
		Min	1
		Mean	1.05

The dataset consists of 6,554,984 records of 54,269 consumers, 2,688 products, and 86 brands, where 6,982 purchases were made by 6,202 consumers, and covering 529 purchased products from 52 brands. The average number of products for each brand is 31.26 with a maximum of 198. The average number of purchases by a consumer is 1.13, with a maximum of 17. The average number of products purchased by a consumer is 1.08, with a maximum of 15; the average number of brands purchased by a consumer is 1.05, with a maximum of 8 (Table 1). To reflect the dynamic variability of asymmetric intra- and inter-brand competition, we first determined a suitable duration to define the time-windows. In doing so, the duration should not be set too short [48, 49], otherwise, the selection behavior within a shopping process will be regarded as the next cycle, resulting inaccurate analysis and sparse data, which is not conducive to the analysis of asymmetric intra- and inter-brand competition. The duration should also not be too long either [48, 49], otherwise useful information might be disregarded, resulting in the changing process of asymmetric intra- and inter-brand competition not being fully reflected. Since the change cycle is determined by the shopping process of consumers, the average duration between search and brand selection within the same shopping process is approximately 15 days, and the duration that consumers search for purchases is related to the product category [13]. In line with this, we also find that 85% of consumers in our dataset had shopping cycles of 14 days (see Appendix II). Considering such behavioral characteristics, we choose 2 weeks as the duration of time-windows, which is also in line with the relevant literature [2, 13].

We use a sliding window filter [2, 48, 66], which is a widely used approach in the analysis of dynamic social networks, [1, 2, 48]. Sliding window filters are used to divide continuous data by time-windows of a fixed size, which can be either overlapping or non-overlapping. The use of overlapping sliding time-windows can improve the accuracy of analysis, as they are less prone to missing important events [2, 48]. Considering the periodicity of consumer behavior, as well as the average purchase cycle of 6.17 days in our dataset, we overlap each consecutive period by one week to reflect the continuous process of changes in brand competition and consequently divide the 10 weeks into 9 windows. For each time-window, we construct consideration sets, while to ensure that brands are competitive, we use only data from consumers in products of the same category [1].

Prior studies often use duration analysis to confirm the relationship between brands based on the purchasing and clicking behavior of consumers [24, 98]. We also use these pieces of information from duration analysis to build a network of competitive relationships between brands. For instance, when a consumer clicks on brands B_1, B_2, B_3 and B_4 within a time-window, and purchases brand B_4, this means that B_4 has a competing relationship with B_1, B_2, and B_3. Accordingly, we filter the clicking behavior of consumers with purchase behavior in each time-window. We employ a modular optimization-based heuristic approach to cluster and visualize communities, which outperforms all other known community detection methods in terms of computational time and community quality [9].

In doing so, we constructed the undirected competition network of Figure 3, which consists of 52 nodes representing the brands, and 916 ties representing competing dyads. We calculated that its density is 0.69, revealing that 69% of the brands have a competitive relationship. The average degree of the undirected competition network is 35.23, and the average path length is 1.35, indicating that the brands within a product category are very closely related. These values indicate that the degree of competition for the brands in our study is particularly fierce, compared to prior relevant studies that have used datasets demonstrating that consumers only click on an average of 2.8 brands before purchasing [13]. The attributes of the competition network between brands, therefore, can indirectly explain the strong competitive relationship among the selected brands in our study.

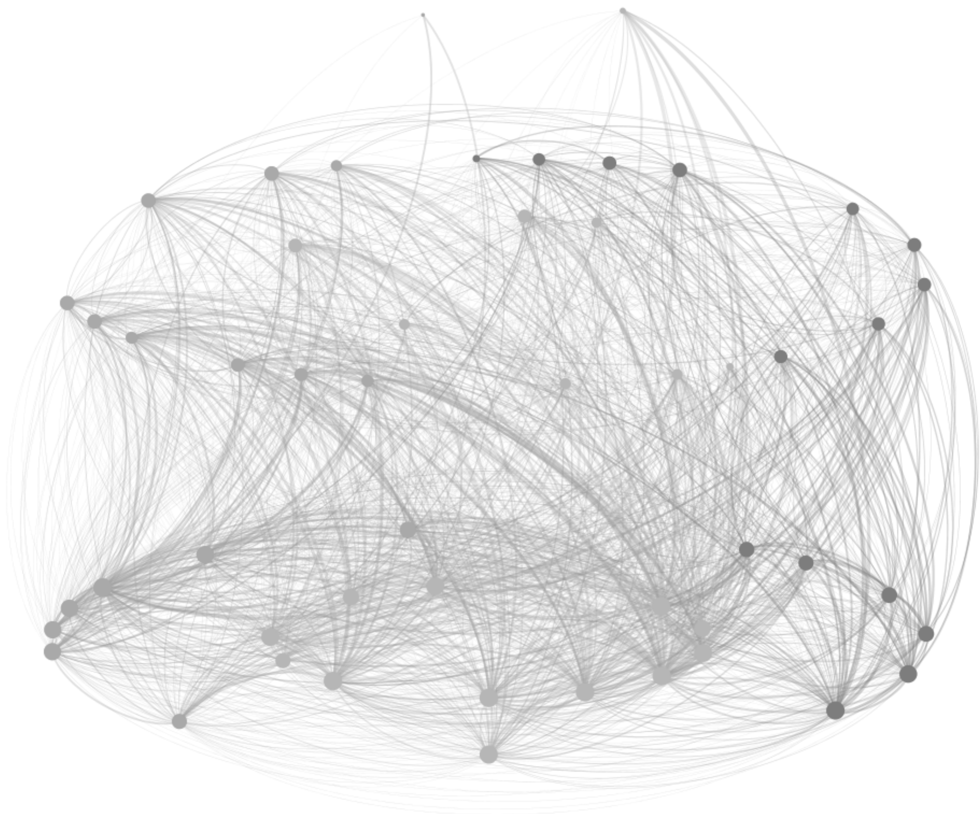

Figure 3: *Network graph of Competition between Any Two Brands*

Note: The nodes represent 52 brands, and their size indicates sales volume. The ties indicate a competitive relationship between brands; the color of the ties represents the market segments. The color version of the figure is provided in Appendix III, Figure II.

Constructing Consideration Sets

After identifying the competing brands, we establish consideration sets in different time-windows based on individual consumers (Figure 2, second step), and then a joint consideration set is constructed for each brand (Figure 2, third step). The joint consideration set is a combination of all consumers in units of brands, and the cell of the joint consideration set is the number of consumers who consider those brands at the same time-window. Specifically, in the matrix, the cell N_{jkt} represents at time-window t, the number of consumers who consider the brands j and k together, and its mathematical expression is shown in the following formula (1):

$$N_{jkt} = \sum_{i \in I} L_{ijt} \times L_{ikt}, (1)$$

where I represents the set of consumers who click on the brands j and k together in time-window t, and I represents one particular consumer. If the consideration set of consumer i includes brand j in the time-window t, then we have L_{ijt}=1; otherwise L_{ijt}=0. Similarly, if the brand k is included in the brand consideration set of the consumer i in time-window t, then we have L_{ikt}=1; otherwise L_{ikt}=0.

Establishing Asymmetric Binary Matrix

The degree of competition between any two brands is asymmetric [75], but this cannot be reflected in formula (1). To illustrate the construction of an asymmetric relations matrix, we take as an example the behavior of the three imaginary consumers presented in Figure 2. Let us assume that during a shopping process, consumer i_1 clicks consecutively on brands B_1 and B_2, consumer i_2 clicks on B_1, B_2 and B_3, and consumer i_3 clicks on B_1, B_2 and B_3. We present these relations in Tables 2, 3, and 4. According to the values in the 4th and 6th columns of Table 2, there is a difference between the consideration sizes. To reflect this difference, we use formula (2) to transform a symmetric matrix into an asymmetric one:

$$N^*_{jkt} = \frac{\sum_{i \in I} L_{ijt} \times L_{ikt}}{\sum_{i \in I} L_{ikt}}, (2)$$

where N^*_{jkt} is the degree of relations between brands j and k in time-window t and the other parameters are as in formula (1).

Table 2: *Constructing Individual Consideration Sets*

Consumer	Click	Consideration set	Consideration set size	Brand	Considered size
i_1	B_1, B_2	B_1, B_2	2	B_1	3
i_2	B_1, B_2, B_3	B_1, B_2, B_3	3	B_2	3
i_3	B_1, B_2, B_3	B_1, B_2, B_3	3	B_3	2
	Mean		2.67	Mean	2.67

Note: The column of consideration sets indicates the number of brands considered by consumers. The column of considered size represents the number of consumers the brand is considered by.

Table 3: *Constructing Joint Symmetric Consideration Sets*

Symmetric consideration sets	$\boldsymbol{B_1}$	$\boldsymbol{B_2}$	$\boldsymbol{B_3}$
B_1	-	3	2
B_2	3	-	2
B_3	2	2	-

Note: The columns represent the number of consumers that consider both brands at the same time.

Table 4: *Constructing Joint Asymmetric Consideration Sets*

asymmetric consideration sets	$\boldsymbol{B_1}$	$\boldsymbol{B_2}$	$\boldsymbol{B_3}$
$\boldsymbol{B_1}$	-	1	0.67
$\boldsymbol{B_2}$	1	-	0.67
$\boldsymbol{B_3}$	1	1	-

Note: Values are symmetrical along the diagonal. To convert them into an asymmetric relationship, it is necessary to divide the values in each row in Table 3 by the values corresponding to each row in the sixth column in Table 2. For example, the number 1 in row 2, column 3 in Table 4 is equal to the value 3 in the corresponding position in Table 3 divided by the value 3 in row 2, column 6 in table 2.

Visualizing Asymmetric Competition

We use Louvain community detection [33] with resolution parameters to cluster the competition relations among brands and products, and we use the in-degree centrality to indicate the degree of competition in the first time-window. Whilst it is possible to form 9 such network graphs, one for each time-window of our study, for brevity we present the first time-window in Figure 4. Each node in Figure 4 represents a brand, and its size indicates sales volume. Colors indicate market segment, while ties indicate that brands are in a consumer's consideration set within a time-window. We use two-way ties to indicate asymmetric competition: the degree of competition for B_1 to B_2 is indicated by the edge weight of B_2 to B_1; conversely, the degree of competition for B_2 to B_1 is indicated by the edge weight of B_1 to B_2, the weight of ties indicates the degree of competition. In different time windows, the number of any two brands considered by consumers at the same time is inconsistent. When any two brands are not considered by any consumers, they are not competitive and do not appear connected.

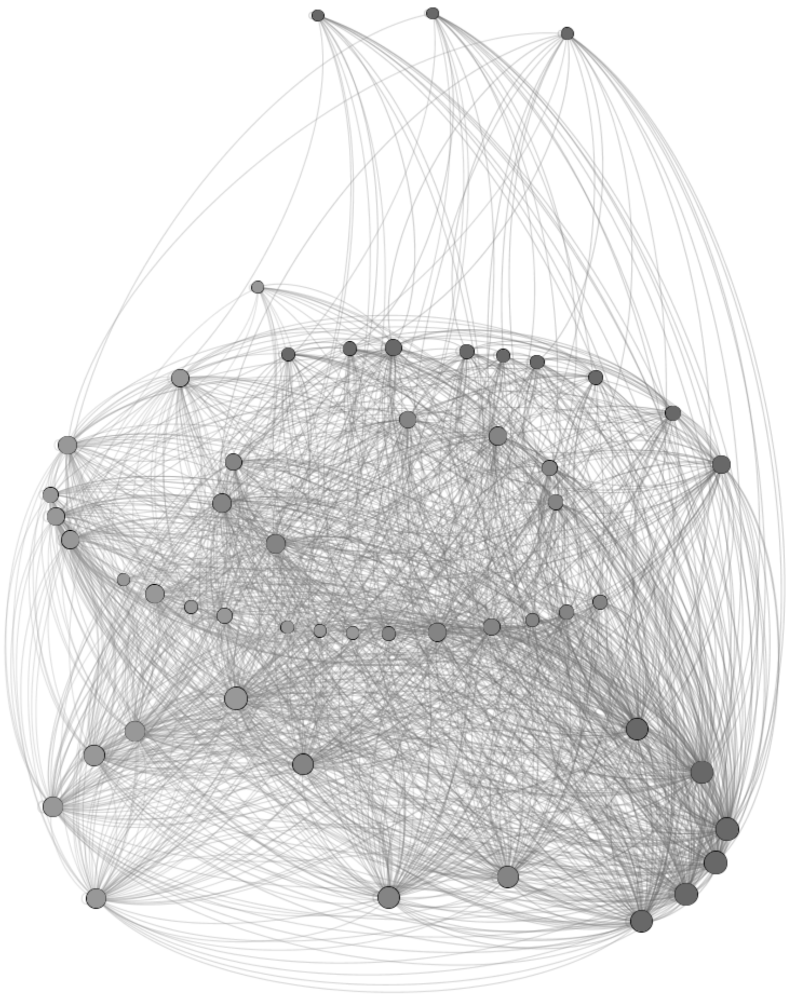

Figure 4: *Network graph of Asymmetric Competition between Brands in the First Time-window*

Note: Each node represents a brand, and their size indicates sales volume. The ties indicate a competitive relationship between brands; the color of the ties represents the market segments. The color version of the figure is provided in Appendix III, Figure III.

INTRA-BRAND CANNIBALIZATION ANALYSIS

By observing the asymmetric competition network in the 9 time-windows, we find that the degree of competition fluctuates, which can pose risks to the operations and market positioning of brands. Therefore, it is necessary to conduct a further in-depth analysis of the underlying reasons for such fluctuations. To further explore such root causes, we take the products of brands as a unit of analysis and construct asymmetric competition networks amongst them, similar to the process for the asymmetric competition networks amongst brands, as we describe below.

Dynamic Competition Process among Brands

To better portray the dynamics of competition amongst brands, we first identify the top 18 brands that account for 90% of the overall market share (see Appendix IV), and then calculate their degree of competition over time, as we demonstrate in Figure 5. We use in-degree centrality in the inter-brand network to indicate the degree of competition. Although the degree of competition fluctuates within time-windows, the fluctuation ranges and can be divided into two phases; the fluctuation range of most brands is within five points of scale and a small number of brands have larger fluctuations (see details within the dashed box in Figure 5). Taking the brand with code '165' as an example, in Figure 5 we can observe that in the third time-window its degree of competition is near 40, while in the fifth time-window its degree of competition is below 5. When the degree of competition fluctuates greatly, it can cause risks to the operations of a brand, which brings challenges to its production and market positioning. There is merit, thus, in further exploring the reasons behind such fluctuations over time.

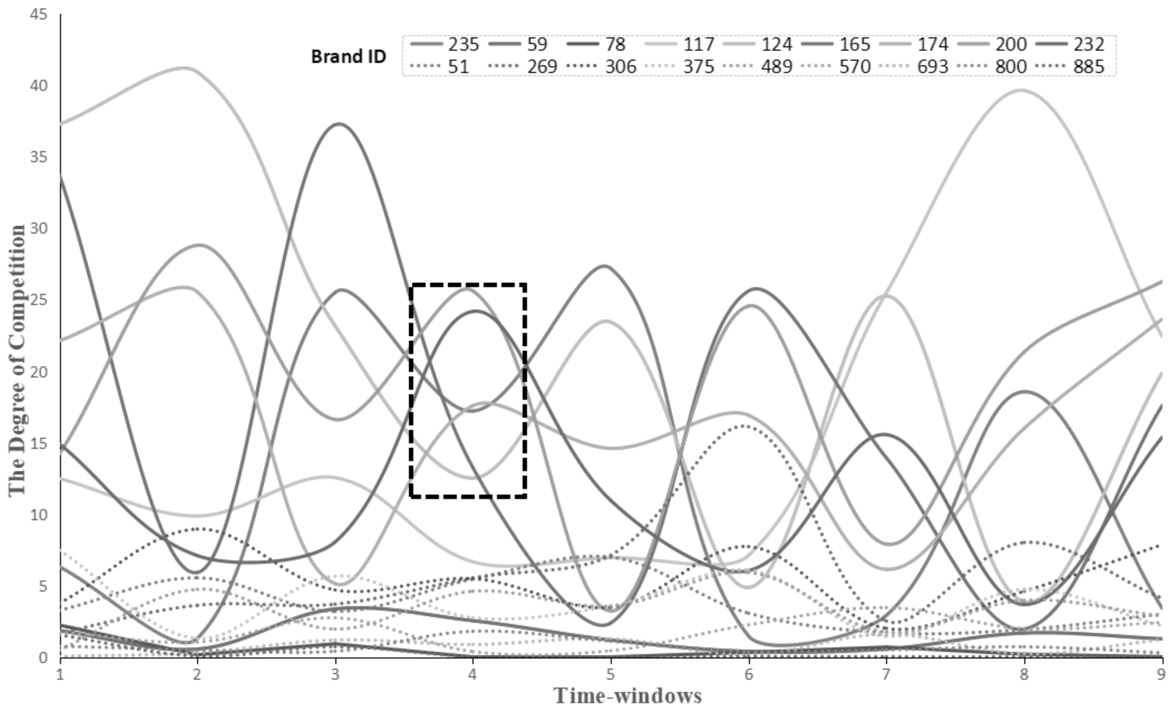

Figure 4: *The Degree of Competition of Each Brand in the Period of Nine Time-windows*

Note: The different colors represent different brands as indicated in the upper right corner. The color version of the figure is provided in Appendix V, Figure V.

Reasons for Dynamic Competition

In the context of online marketplaces, 10% of products account for 90% of the market share of a brand [90, 98], similar to the Pareto principle [14, 86]. We, thus, select the products with the top 90% sales volume for each brand and analyze the intra- and inter-brand competition in different time-windows.

The results are shown in Figure 6, which consists of five directed network graphs that represent intra- and inter-brand competition in different time-windows. Each node represents a product, the numbers in the first graph represent the product code, and the products represented by the nodes in the subsequent graphs are the same as the first one. Nodes of the same color belong to the same brand, as shown below Figure 6a. Bilateral ties show the existence of intra- and inter-brand competition, while unilateral ties represent that one product poses a competitive threat to the others, but the reverse is not true. The weight of ties represents the degree of intra- and inter-brand competition. The number of ties between products within a brand and their weight is significantly higher than that between brands. This shows that intra-brand competition is much higher than inter-brand one and that consumers tend to consider products within the same brand rather than across different brands. By examining the competition network in different time windows, we can see that the ties and their thickness change over time, which demonstrates that the degree of intra- and inter-brand competition changes dynamically.

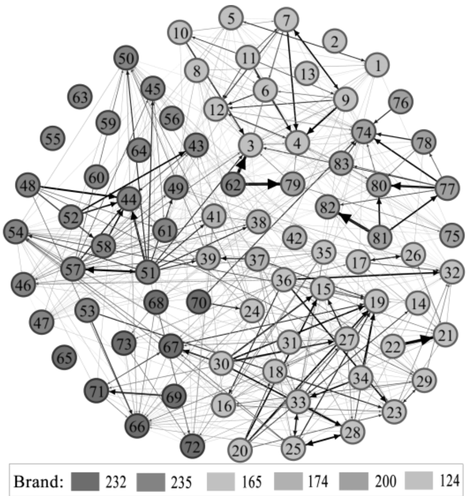

Figure 6a: *First Time-window*

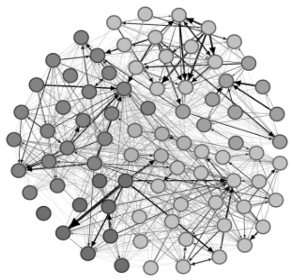

Figure 6b: *Third Time-window*

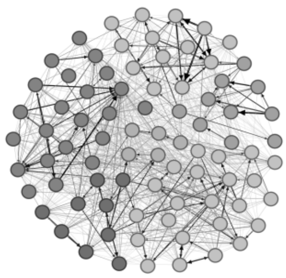

Figure 6c: *Fifth Time-window*

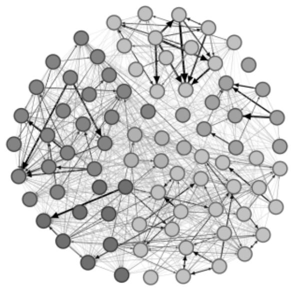

Figure 6d: *Seventh Time-window*

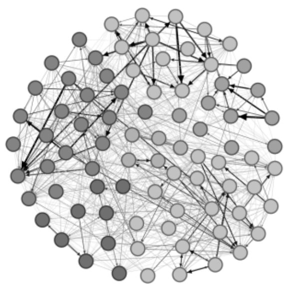

Figure 6e: *Ninth Time-window*

Figure 6: *Intra- and Inter-brand competition in different time-windows. The color version of the figure is provided in Appendix VI, Figure VI.*

To verify the above inference, we divided the 52 brands into three categories based on their sales volume (high, medium, low), and randomly selected one brand from each category. The intra-brand relations among products are divided into 9 asymmetric networks. We calculate the indicators of each product and analyze the relationship between the indicators and the sales volume with a mixed regression. We find that 83.9% of sales fluctuations are contributed by intra-brand competition

(Appendix VII, Table II). The main reason for intra-brand competition is the high substitutability of products. Thus, optimizing product offerings is conducive to enhancing the degree of competition between brands. We can, thus, conclude that the main reason for fluctuations in the degree of brand competition is the interaction among products within the brand. The analysis of six brands with larger fluctuations of competitive relation shows that intra-brand competition is more intense than intra-brand one, which is in line with the extant literature on the topic [52, 75]. We can, thus, infer that internal production-line adjustments are needed to optimize offerings within the brand, and we can calculate in advance how the market share of related products will change after such an adjustment.

Reduce Intra-Brand Cannibalization

Product proliferation can result in an increase in production cost, and consequently a decrease in profit margins. Brands, therefore, need to decide whether to adjust the original production line and estimate the loss/gain of market share when doing so. In the context of online marketplaces, the sales distribution of brands is considered fat-tailed [11, 57]. Based on the asymmetric networks, we can model sales after adjusting the production line. Subsequently, we randomly select the brand with code '235' as an example to illustrate the market-share redistribution when its production line is adjusted.

As we depict in Table 5, before production line adjustment, the market share of this brand in the first time-window was 29.09%, of which the market share of the top (i.e., best-seller) six products was 23.77%, and of the bottom (i.e., least sold) ten products was 5.32%. To further illustrate, let us assume the possibility of stopping producing the products with sales volume ranking of the bottom ten products for the brand with code '235', and also assume that in the consumers' consideration set each product has an equal probability of being considered. If consumer A considers 4 products at the same time-window, the probability of each product being considered is 25%, while if the number of products in the market decreases by one, the possibility of consumers considering other products increases. We use the consumers' consideration set, by considering the idea that the probability of each product being considered is equal and we calculate the adjustment of market share after the brand reduces the production of the bottom ten products for the brand with code '235'. According to the result, we find that the adjusted market share does not decrease by 5.32%, but only by 3.08%. To the extreme, assuming that the production cost of each product is equal, producing six products only requires 37.5% of the overall production cost for 16 products, and hence removing the bottom ten products can save 62.5% of the original production costs. Thus, the inter- and intra-brand asymmetric competition network can also provide a strong reference for the production line adjustment.

Table 5: *Market-share Distribution with Product-line Adjustment*

Changing the overall market share of brand #235 before and after product adjustment				
	Brand ID	**Old share (%)**	**New share (%)**	**Share gain (%)**
Top six products	235	23.77	26.01	2.24
Bottom ten products	235	5.32	0.00	−5.32
All	235	29.09	26.01	−3.08

Products with the largest share changes after the product adjustment				
Brand ID	**Product ID**	**Old share (%)**	**New share (%)**	**Share gain (%)**
235	44	3.84	4.53	0.69
235	46	3.08	3.57	0.49
235	45	4.00	4.43	0.43
235	43	1.99	2.35	0.36
232	66	0.35	0.64	0.29

INTER-BRAND COMPETITION ANALYSIS

The degree of brand competition is a manifestation of market power, and it can be measured through the sales volume [79]. To capture competition dynamics between brands, we use the frequency of consumers' conditional considerations to measure the degree of competition [75]; the process is shown in Table 3 and Table 4. According to network theory [5, 73, 76], different structural characteristics represent different consumers' consideration sets, indicating that preference for a brand varies, having the potential to model sales. In line with this, therefore, we use the span of structural holes, betweenness centrality, and degree centrality as IV to model sales. To solve the auto-correlation problem, we use spatial auto-regressive (SAR) models for parameter estimation [95].

Variable Selection

We consider the frequency of consumers' conditional considerations as directly proportional to their actual purchase [64] and use brand sales to measure the degree of competition. Due to the significant difference in sales volume of brands, we take the *log* of sales $logy_{jt}$ as DV, where j indicates a brand and t the time-window. We use structural constraint indicators [23, 93], and for each time-window the indicators are estimated by formulas (3) and (4):

$$C_{jt} = \sum_{k} C_{jkt}, (3)$$

$$C_{jkt} = \left(p_{jkt} + \sum_{q} p_{jqt} p_{qkt} \right)^2, (4)$$

where C_{jt} represents the network constraint of brand j in the time t, where the larger the value, the greater the constraint and the fewer structural holes a brand would have. Since structural constraint is negatively related to the degree of structure holes, we use $1 - Constraint$ to represent structural holes index. K represents a brand directly adjacent to j, and C_{jkt} indicates the degree of constraint of brand k directly adjacent to brand j at time t. q represents a common adjacent point for brand j and k. p_{jkt} represents the proportion of brand j in the neighboring nodes of brand k at time t. $\sum_q p_{jqt} p_{qkt}$ indicates the indirect proportion of brand k on brand j. Considering that the network is asymmetric, p_{jkt} is not equal to p_{kjt}. Betweenness centrality is calculated according to formula (5) [34]:

$$C_B(j_t) = \frac{g_{klt}(j)}{g_{klt}}, (5)$$

where g_{klt} is the number of shortest paths between nodes k and l in time t, and $g_{klt}(j)$ represents the number of shortest paths through node j. Therefore, formula (5) shows how often node j is located on the shortest path between nodes k and l in time t. The formula for degree centrality is [34, 35]:

$$D_{jt} = C_D(j) \sum_{K}^{N} x_{jkt}, (6)$$

where j is a focal node, k is a node other than j. N is the number of all nodes; x is the adjacency matrix, and the cell x_{jkt} indicates the direct adjacency between nodes j and k in time t. When the value is 1, there is adjacency between nodes; otherwise, the value is 0. Degree and betweenness centrality measure the importance of a structural position of a node in a network; specifically, degree centrality measures the local influence of a node, while betweenness centrality measures the global impact of a node [92]. In Table 6 we summarize the descriptive statistics of the network indicators in our study.

Table 6: *Descriptive Statistics of Variables*

Variables		Sample size	Mean	St. dev	Min	Max
DV	**Log-sales**	468(52*9)	1.45	1.48	0.00	5.33
IV	**1-Constrain**	468(52*9)	0.91	0.17	0.71	0.93
	Degree	468(52*9)	71.68	5.33	15.00	78.00
	Betweenness	468(52*9)	28.03	27.40	0.00	175.04
CV	**Gender**	468(52*9)	0.09	0.20	0.00	1.00
	Membership level	468(52*9)	0.12	0.19	0.00	0.87
	Age	468(52*9)	0.37	0.31	0.00	1.00

Control Variables

Since the intra-brand asymmetric competition network is constructed by consideration sets on the brand level, the demographics of consumers are considered to have an impact on sales [55]. We, thus, control for the effects of gender, age, and membership level of consumers in our models. As females tend to have a higher purchase intention and a more positive attitude towards brands [81], we control for gender on sales using the percentage of female in the dataset. Age is also a key factor of individual motivation [29], as consumers of different ages have different purchasing motives, and options.

The consumers that completed a purchase in our dataset are between 26 and 45 years old. Therefore, we use the median value of 36 years and the percentage of consumers above 36 years of age to analyze the impact of age on the volume of sales. The membership level can explain the purchase and consumption experience of consumers, as the higher the membership level, the richer their shopping experience. Consumers with different experiences have different clicking behaviors and purchase possibilities. Thus, the membership level can impact brands sales. There are four levels of memberships in our dataset, according to the descriptive statistics, most of the purchasers are above level 4. Thus, we use the percentage of consumers on level 4 in our dataset to measure the effect of membership level. The correlations for all variables of interest are presented in Appendix VIII (Table III).

Estimation and Results

An extreme strategy in parameter estimation is to treat panel data as cross-sectional—since there is no individual effect—and perform pooled regression. We use a pooled regression model as baseline and a population-averaged (PA) estimator for the parameters. To avoid potential endogeneity issues, we lag the IV by one period [7, 10, 74]. Considering the effects of IV and CV on the *log* of sales volume, our model is shown in formula (7), in which u_j is the influence of the brand, and ε_{jt} is the error term:

$$logy_{jt} = logsale_{jt} = \alpha + \beta_0\left(1 - Constraint_{jt-1}\right) + \beta_1 BetweennessCentrality_{jt-1} + \beta_2 DegreeCentrality_{jt-1} + \beta_3 Gender_{jt-1} + \beta_4 MembershipLevel_{jt-1} + \beta_5 Age_{jt-1} + \mu_j + \xi_{jt}, \quad (7)$$

The estimation results of the baseline model are shown in Table 8 (models 1 and 2). To test multicollinearity among IV and CV, we performed a variance inflation factor (VIF) test [80]. We present the results in Appendix IX, where the maximum value of VIF for all variable is 2.01, far less than its threshold of 10 [80], indicating that multicollinearity among IV and CV is not an issue. However, the basic assumption of using OLS model for parameter estimation is that there is no auto-correlation among DV. Since the ties in the inter-brand network are derived from a clicking behavior in which consumers can be the same or overlapping, an auto-correlation between the individuals of a brand in different time-windows is possible. We use the global spatial auto-correlation Moran detection [53] to test for auto-correlation among DV. According to the results in Table 7, the Moran index among the DV is -0.096, and the *p* value shows that the exponential relationship is significant, indicating a strict auto-correlation between DV. The negative value of the Moran index indicates a negative correlation in the sales of each brand, which is consistent with their competitive relationship.

Table 7: *Result of Global Spatial Auto-correlation Test*

Variable	Moran index	E (index)	SD (index)	z	p
Log sales	-0.096	-0.002	0.011	-8.494	0.000

Note: Moran index is a comprehensive evaluation used to measure the degree of spatial auto-correlation; E (index) represents the mean of Moran index; SD (index) represents the variance of Moran index; z represents the test method, when the z value is positive and significant, it indicates that there is a positive spatial auto-correlation, when the z value is negative and significant, it indicates that there is a negative spatial auto-correlation; The p value is the result of a two-sided test.

SAR models come from spatial econometrics and have recently been widely adopted in neighboring management disciplines, mainly due to the increasing attention paid to the interaction between economic factors, such as peer effect, neighborhood effect, and spillover effect, as well as the rapid development of geographic IS, and the wide availability of spatial data. Whilst SAR was originally introduced for spatial data [6], it has been increasingly gaining popularity in all variances of network analysis [22, 26, 56], as well as for modeling interdependent consumer preferences [94], since network data share many similarities with spatial data. For our study, the main reason for the existence of auto-correlation between the sales of each brand is that the ties in the inter-brand competition network are interdependent [73]. Any two different brands in our dataset are related to each other through a series of shared consumers. For each brand in the same product category, since the total number of consumers is fixed, the overall market demand remains unchanged, and there is a negative correlation between the sales of each brand. Since there are three types of SAR models, we detail our selection in Appendix X, considering DV as shown in formula (8):

$$SAR: Y_{jt} = \lambda \sum_{k=1}^{N} W_{jk} Y_{kt} + X_{jt-1}\beta + \mu_j + \delta_t + \epsilon_{jt}, (8)$$

where the parameters are consistent with the previous statements. W_{jk} is the 52x52 dimensional spatial weight matrix, in which cell values are the mean of all the brand asymmetric matrices. Y_{kt} indicates the sales volume of other brands at the same time-window. X_{jt-1} is the IV and CV of brand j lagging one period. Δ_t is the time effect, referring to the nine time-windows. The estimations of Model 3 and 4 in Table 8 are the corresponding results of the SAR model, and the influences of the three network measures on sales are significant. In addition, among the CV, only gender has a significant impact on sales. Meanwhile, according to the value of R square, we can conclude that the SAR model has a better fitting effect on the actual sales (0.687 > 0.539) compared to the baseline model.

Table 8: *Estimation Results for OLS and SAR (DV=Log Sales)*

Variables		**Model 1**	**Model 2**	**Model 3**	**Model 4**
		OLS	**OLS**	**SAR**	**SAR**
CV	**Gender**	0.998*** (0.323)	0.879*** (0.246)	0.395** (0.166)	0.393** (0.160)
	Membership level	3.259*** (0.331)	2.040*** (0.262)	0.165 (0.185)	0.185 (0.175)
	Age	0.308 (0.201)	0.323** (0.153)	0.07 (0.187)	0.083 (0.184)
IV	**1-Constrain**		-53.905*** (3.788)		-4.244* (2.337)
	Degree Centrality		0.157*** (0.012)		0.017** (0.007)
	Betweenness Centrality		0.022*** (0.002)		0.002** (0.001)
Network	**Rho**			0.602*** (0.06)	0.589*** (0.063)
	Constant	0.850*** (0.115)	38.099*** (3.032)	0.779*** (0.197)	3.306* (1.721)
R square		**0.200**	**0.539**	**0.160**	**0.687**
AIC				**794.026**	**791.847**

*Note: Standard errors are in parentheses. Path coefficients are outside parentheses. Rho represents the correlation of a competitive relation between a neighbor and a focal brand. R square represents the fitting effect of the corresponding model on actual sales. AIC represents the goodness-of-fit of the SAR model, and the lower the value the better the goodness-of-fit. * p<0.1; ** 0.01<p<0.05; *** p<0.01.*

Robustness Check

Different product categories have different attributes, and the behavior of consumers varies per category [50]. To test the robustness of our results, we use the clicking and purchasing behavior in different product categories. In Table 9, we compare the structures of the original and the test dataset. Compared to the original product category data, the product category selected in the test-set differs in three ways: i) the number of purchased brands (in the test-set is about half of the original); ii) the number of consumers with click and purchase records (it is double, while the number of clicks is roughly equal); and iii) the purchase amount of the test-set (it is nearly double that of the original).

Table 9: *Comparison of the Original Dataset and the Test Dataset*

	Original dataset	Test dataset
The number of clicks	6554984	6787974
Purchases	6982	13281
Users	54405	104740
Purchasers	6202	12921
Products	2688	3938
Purchased Products	529	859
Brands	86	40
Purchased Brands	52	30

The comparative analysis shows that the number of clicks on each product in the original dataset is almost twice that of the test-set, which indicates that brand's attributes of the two types are considerably different, and there is a large difference in the identification process of consumers before the purchase. Concurrently, in the test-set, the average purchase amount of each consumer is 1.3 times that of the original dataset, indicating that products in the test-set are more aligned with the characteristics of daily consumables than the products in the original dataset. The comparison of the two data structures shows that the two products categories have greater heterogeneity. Heterogeneity refers to the degree of variance in the attributes of the focal network [61] and captures the degree of differences in consumers' browsing records, repurchases, as well as their consumption frequency.

Since the product characteristics of the two datasets are different, to reduce the dimensionality impact we standardized dataset indexes before comparing the product network. If the test-set results are consistent with the original, our findings can be considered robust. Thus, we re-analyze the test dataset according to the above process, and the results (Appendix XI, Table VII) show that the three network measures affect the brand sales in the same way as the original dataset. Specifically, the results of all IV are significant, and structural holes have a negative impact on brand sales, while degree centrality and betweenness centrality have positive impacts on brand sales, thus successfully demonstrating the robustness of our findings. Further, prior research shows that consumers' consideration sets are influenced by weekends and weekdays [8]. Thus, to further verify the robustness of our findings, we divide consumers' weekly consideration sets in the test-set into two-time windows: weekend and weekdays. The transformation process between consumers' consideration sets and brand dynamic competition network in different time windows, and the analysis of the influencing factors of brand dynamic degree of competition are consistent with the analysis of the original data. The results (Appendix XI, Tables VII-XI), show that structural holes have a negative impact on sales, in-degree centrality has a positive impact on sales, and betweenness centrality has no significant effect on sales. These results further verify the robustness of our findings.

Endogeneity Test

Whilst the lagging of variables by one period can effectively alleviate the endogeneity issues of mutual causality, we still account for the two-way relation between sales of different brands within a time-window. We use the lagging *log* sales volume of competing brands within a time-window as an instrumental variable because the previous sales of competing brands affect their current sales, and the sales of competing brands at the same time-window affect each other. To further analyze the influence of the network measures on the volume of sales, we use a 2-Stage Least Square (2SLS) regression to address the endogeneity issues in line with Perera and Tan [71]. We follow the same approach for parameter analysis, showing the results of the endogeneity test in Table 10. Subsequently, we carried out a Sargan's Test, showing that the instrumental variable is not used excessively (p = .94). The influence of network measures on sales is still consistent with the SAR model and, therefore, suggests that endogeneity is not an issue in our analysis.

Table 10: *Endogeneity Test (DV=Log Sales)*

Variables		2SLS Model	SAR Model
CV	**Gender**	0.446*** (0.166)	0.393** (0.160)
	Membership level	0.423** (0.185)	0.185 (0.175)
	Age	0.135 (0.187)	0.083 (0.184)
IV	**1-Constrain**	-7.370*** (2.564)	-4.244* (2.337)
	Degree Centrality	0.025*** (0.008)	0.017** (0.007)
	Betweenness Centrality	0.004*** (0.001)	0.002** (0.001)
Network	**Rho**		0.589*** (0.063)
	Constant	4.925** (1.927)	3.306* (1.721)
R square			**0.687**
AIC			**791.847**

*Note: Standard errors are listed in parentheses. Path coefficients are listed outside the parentheses. Rho represents the correlation of competitive relations between a neighbor brand and a focal brand. R square represents the fitting effect of the corresponding model on actual sales. AIC represents the goodness-of-fit of the spatial auto-regressive model, and the lower the value the better the goodness-of-fit. * p<0.1; ** 0.01<p<0.05; *** p<0.01.*

DISCUSSION

Key Findings

Research on brand competition in the context of online marketplaces from the perspective of product offerings has timely reference value [30, 52, 75]. Prior studies have identified the market segments and *static* asymmetric competition relations of brands by using consumers' consideration sets. Brands on online marketplaces, however, compete for consumers' attention in real-time, and consequently, being able to identify asymmetric competition in such a context is more challenging than in traditional ones. As the preferences and attention span of consumers constantly change, capturing their clickstream data and dynamically modeling the degree of competition can effectively reduce the potential risks for brands of losing market share. We propose a novel approach to identify brand product cannibalization and market competition relations through consumers' consideration sets and use the volume of sales to verify this. Subsequently, we first analyze how brands should adjust their production lines to reduce intra-brand cannibalization. We further explore the factors that affect future sales of brands from the perspective of inter-brand competition. We use the span of structural holes, betweenness centrality, and degree centrality to analyze the volume of sales. Finally, regarding future sales, we handle the interdependence among brands, using SAR models for parameter estimation.

In doing so, our study brings forward three key findings. First, we find that intra-brand competition has a greater impact on the degree of dynamic competition than inter-brand. Considering the importance of intra- and inter-brand competition in optimizing product offerings, our findings underscore the need for brands to balance these two effects when offering products on online marketplaces. If the impact of intra-brand competition on the degree of dynamic competition is more pronounced, brands on online marketplaces are better off when reducing the number of their offerings, which addresses the research question of our study. This finding may be counterintuitive, as in practice online marketplaces are not limited by capacity or physical shelf space. In addition, our work can also predict in advance how the market share will change after product line adjustment.

Second, our findings show that products residing in structural holes serve as an outbound agent and hence create negative effects on sales volume, which broadly contradicts the insights from the extant management literature on the positive effects of the span of structural holes on organizational performance. When a brand is in a structural hole, this indicates that fewer consumers consider it at the same time along with other brands. For such brands, however, the degree of competition in a product category is relatively weak. Therefore, our findings demonstrate that in the context of online marketplaces, brands should minimize heterogeneity in their product offerings within their market segment to increase sales, which also contradicts intuitive current practices for pursuing individualization in product offerings to become more competitive [28].

Third, our findings show the positive impact of degree centrality and betweenness centrality on the volume of brand sales on online marketplaces, which validates the two initial assumptions of our study. We find that the influence of degree centrality is higher than betweenness centrality, which means that intra-brand cannibalization is more intense than inter-brand competition. Thus, when brands consider their product design and marketing strategies for online marketplaces, they should give a higher reference weight to the characteristics of each brand in their market segments.

Theoretical Implications

Our work brings forward four key theoretical contributions. First, we consider an asymmetric competition relationship amongst brands, which is more consistent with contemporary reality, and provides better theoretical foundations for competition on online marketplaces. In addition, we balance intra-brand cannibalization and inter-brand competition to reduce the market risk of product offerings in this context. Second, we use consumers' time-variant implicit preferences to unearth the *dynamics* of intra- and inter-brand asymmetric competition and provide a theoretical reference on *when* and *how* brands should optimize their offerings on online marketplaces to increase their sales volume. Further to these, we integrate demand information of consumers' implicit preferences with supply-side sales insights to increase the robustness of our findings. By doing so, we extend the literature on the use of consumers' time-variant implicit preference information from the consumer level [41, 60] to the brand level. Online marketplaces, due to their high product aggregation and fast updates on offerings, can result in constantly changing consumers' preferences and attention spans. Capturing the *dynamics* of intra- and inter-brand asymmetric competition can provide a sorely needed theoretical reference on *when* and *how* brands should optimize their offerings on online marketplaces, thereby reducing the market risk. Third, we combine the span of structural holes with centrality measures to model the sales of brands on online marketplaces. In doing so, we evaluate the effects of structural holes and centrality measures and distinguish their impact on sales from the local and global markets. In addition, we expound the attributes of the brand competition network from the perspective of consumers' consideration sets, expanding the applications of network theory to capture brand competition. Finally, we use consumers' behavioral information reflecting their implicit preferences to consider the dynamics of asymmetric competition. In doing so, we explore the competition between brands and demonstrate the complex dynamics of product cannibalization. Our work, thus, shifts the discussions on asymmetric competition from a *static* to a *dynamic* paradigm. Consequently, the findings of our study provide a novel theoretical standpoint for brands on online marketplaces to predict sales volume based on their position in the global and local markets, which ultimately can provide information reference for product design as well as marketing strategies in this context.

Practical Implications

Our work also provides valuable insights for practice. Our findings demonstrate that in the context of online marketplaces the degree of inter-brand product cannibalization is higher than the degree of external competition among brands. Therefore, brands operating in this context need to adjust the length of the production line to optimize their offerings and increase their sales. Our findings, thus, suggest that to enhance the degree of competition on online marketplaces, brands must adjust their product line in ways fundamentally different from the contemporary industry practices. To this end, our work can provide brands with insights for predicting the redistribution of market shares, identify market risks, improve decision-making, and ultimately increase their sales. We show that to increase sales, brands should increase the number of ties with others in their asymmetric competition network and minimize heterogeneity in product offerings within their market segment. In practical terms, this finding means that brands should identify products with similar attributes to those with high sales in their target market segment when offering products on online marketplaces. Consumables and newly entered products usually take sales volume as the main goal, therefore, brands should consider the homogeneity of related products in the market segment, since most of online consumers have a lower level of perceived differences in such products [75]. Our work, therefore, provides brands that operate on online marketplaces with novel insights that contradict contemporary practices. Our practical insights, thus, provide a bedrock for novel business models and enable brands to increase their sales.

Limitations and Future Research

Whilst we followed a thorough and structured research design, there are limitations that we need to acknowledge, which concurrently open future research avenues. The clicking behavior of consumers includes actions such as browsing, clicking, adding, or removing goods to the shopping cart, as well as purchasing. Whilst such actions might stand for the preferences of consumers towards a brand, they share a uniform weight in our estimations. Future research, therefore, could assign different weights to the various actions of consumers, and reflect on the strength of sales. Moreover, in our study, we only consider the spatial auto-correlation problem when using network measures to reflect the future sales of brands. Due to the cumulative effect of brand sales, however, there is also a possible temporal correlation problem, and we encourage future research endeavors to reflect on the possibility of such an issue. Third, we only consider the volume of sales as a measure of brand competition, but we do not consider the impact of profit on the degree of competition. Our conclusions, thus, may be different from other profit-oriented studies, and future research endeavors could use brand sales along with profit for comparative analysis according to the strategic goals of brands. Further to these, when analyzing the dataset of our study, we considered the sequence of consumers' behaviors in different time-windows, but we did not consider the sequence of their clicking behaviors within the same time-window, as this was beyond the scope of our paper. In addition, all our conclusions are based on the analysis of empirical data, and no additional experiments have been conducted. We encourage future research endeavors, therefore, to consider the sequence of consumers' consideration sets within the same time-window when exploring the competition amongst brands, as well as to consider conducting field experiments. Finally, our work highlights the importance of access to information for extracting actionable insights, which makes paramount the quality of such information [82, 84], especially for datasets of high volume, variety, and veracity [3, 4, 45, 63]. Future research endeavors on the topic, thus, should further contribute to the extant IS research agenda [83] by focusing on the challenges surrounding information quality in the use of large datasets for extracting actionable insights.

CONCLUSIONS

By examining market dynamics and the influence of network positions, we shed light on how brands can balance intra-brand cannibalization and inter-brand competition in their product offerings, ultimately enhancing their degree of competition on online marketplaces. The findings of our study offer a theoretical reference for brands seeking to optimize their offerings on online marketplaces. Specifically, if the impact of intra-brand competition on the degree of dynamic competition is more pronounced, brands would benefit from reducing the number of their offerings on online marketplaces. We further suggest that brands should strive to minimize heterogeneity in their product offerings within their market segment to increase sales. Moreover, brands should assign greater weight to the unique characteristics of each brand within their market segment. By considering these factors, brands can better position themselves to thrive in highly competitive online marketplaces.

REFERENCES

1. Andrade-Rojas, M.G., Kathuria, A., and Konsynski, B.R. Competitive Brokerage: How Information Management Capability And Collaboration Networks Act As Substitutes. *Journal of Management Information Systems*, 38, 3 (2021), 667-703.
2. Angelopoulos, S., and Merali, Y. Sometimes a cigar is not just a cigar: Unfolding the transcendence of boundaries across the digital and physical. *ICIS*, Seoul, Korea: Association for Information Systems, 2017.
3. Angelopoulos, S., Brown, M., McAuley, D., Merali, Y., Mortier, R., and Price, D. Stewardship of personal data on social networking sites. *International Journal of Information Management*, 56 (2021), 102208.
4. Angelopoulos, S., Bendoly, E., Fransoo, J.C., Hoberg, K., Ou, C.X., and Tenhiälä, A. Digital transformation in operations management: Fundamental change through agency reversal. *Journal of Operations Management*, Forthcoming (2023).
5. Angelopoulos, S., Canhilal, K.S., and Hawkins, M.A. From Groups to Communities: A Resource Mobilization Theory Perspective on the Emergence of Communities. *Information Systems Frontiers* (2023), 1-18.
6. Banerjee, S., Carlin, B.P., and Gelfand, A.E. *Hierarchical modeling and analysis for spatial data*. CRC press, 2014.
7. Bellemare, M.F., Masaki, T., and Pepinsky, T.B. Lagged explanatory variables and the estimation of causal effect. *The Journal of Politics*, 79, 3 (2017), 949-963.
8. Bhatnagar, A., Sen, A., and Sinha, A.P. Providing a window of opportunity for converting eStore visitors. *Information Systems Research*, 28, 1 (2017), 22-32.
9. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008, 10 (2008), P10008.
10. Blundell, R., and Bond, S. Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87, 1 (1998), 115-143.
11. Bordley, R. Determining the appropriate depth and breadth of a firm's product portfolio. *Journal of Marketing Research*, 40, 1 (2003), 39-53.
12. Brands, R.A., Menges, J.I., and Kilduff, M. The leader-in-social-network schema: Perceptions of network structure affect gendered attributions of charisma. *Organization Science*, 26, 4 (2015), 1210-1225.
13. Bronnenberg, B.J., Kim, J.B., and Mela, C.F. Zooming in on choice: How do consumers search for cameras online? *Marketing science*, 35, 5 (2016), 693-712.
14. Brynjolfsson, E., Hu, Y., and Smith, M.D. Research commentary—long tails vs. superstars: The effect of information technology on product variety and sales concentration patterns. *Information Systems Research*, 21, 4 (2010), 736-747.
15. Brynjolfsson, E., Hu, Y., and Simester, D. Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management science*, 57, 8 (2011), 1373-1386.

16. Bucklin, R.E., and Sismeiro, C. A model of web site browsing behavior estimated on clickstream data. *Journal of marketing research*, 40, 3 (2003), 249-267.
17. Burt, R.S. The social structure of competition. *Networks in the knowledge economy*, 13, 2 (2003), 57-91.
18. Burt, R.S. Structural holes and good ideas. *American journal of sociology*, 110, 2 (2004), 349-399.
19. Chae, S., Yan, T., and Yang, Y. Supplier innovation value from a buyer–supplier structural equivalence view: Evidence from the PACE awards in the automotive industry. *Journal of Operations Management*, 66, 7-8 (2020), 820-838.
20. Chen, M.-J. Competitor analysis and interfirm rivalry: Toward a theoretical integration. *Academy of management review*, 21, 1 (1996), 100-134.
21. Chen, M.-J., and Miller, D. Competitive dynamics: Themes, trends, and a prospective research platform. *Academy of management annals*, 6, 1 (2012), 135-210.
22. Chen, X., Chen, Y., and Xiao, P. The impact of sampling and network topology on the estimation of social intercorrelations. *Journal of Marketing Research*, 50, 1 (2013), 95-110.
23. Chi, L., Ravichandran, T., and Andrevski, G. Information technology, network structure, and competitive action. *Information systems research*, 21, 3 (2010), 543-570.
24. Chintagunta, P.K., and Haldar, S. Investigating purchase timing behavior in two related product categories. *Journal of Marketing Research*, 35, 1 (1998), 43-53.
25. Clement, J., and Puranam, P. Searching for structure: Formal organization design as a guide to network evolution. *Management Science*, 64, 8 (2018), 3879-3895.
26. Cohen-Cole, E., Liu, X., and Zenou, Y. Multivariate choices and identification of social interactions. *Journal of Applied Econometrics*, 33, 2 (2018), 165-178.
27. DeSarbo, W.S., and Grewal, R. An alternative efficient representation of demand-based competitive asymmetry. *Strategic Management Journal*, 28, 7 (2007), 755-766.
28. Dickson, P.R., and Ginter, J.L. Market segmentation, product differentiation, and marketing strategy. *Journal of marketing*, 51, 2 (1987), 1-10.
29. Diehl, M., and Hay, E.L. Self-concept differentiation and self-concept clarity across adulthood: Associations with age and psychological well-being. *The International Journal of Aging and Human Development*, 73, 2 (2011), 125-152.
30. Dou, W., Lim, K.H., Su, C., Zhou, N., and Cui, N. Brand positioning strategy using search engine marketing. *MIS quarterly* (2010), 261-279.
31. Draganska, M., and Jain, D.C. Product-line length as a competitive tool. *Journal of Economics & Management Strategy*, 14, 1 (2005), 1-28.
32. Duan, W., and Zhang, J. The comparative performance of online referral channels in e-commerce. *Journal of Management Information Systems*, 38, 3 (2021), 828-854.
33. Fortunato, S., and Barthelemy, M. Resolution limit in community detection. *Proceedings of the national academy of sciences*, 104, 1 (2007), 36-41.
34. Freeman, L.C. Centrality in social networks conceptual clarification. *Social Networks*, 1, 3 (1978), 215-239.
35. Freeman, L.C., Roeder, D., and Mulholland, R.R. Centrality in social networks: II. Experimental results. *Social networks*, 2, 2 (1979), 119-141.
36. Funk, R.J., and Owen-Smith, J. A dynamic network measure of technological change. *Management science*, 63, 3 (2017), 791-817.
37. Gelper, S., van der Lans, R., and van Bruggen, G. Competition for attention in online social networks: Implications for seeding strategies. *Management Science*, 67, 2 (2021), 1026-1047.
38. Giachetti, C., and Dagnino, G.B. Detecting the relationship between competitive intensity and firm product line length: Evidence from the worldwide mobile phone industry. *Strategic Management Journal*, 35, 9 (2014), 1398-1409.
39. Gnyawali, D.R., Fan, W., and Penner, J. Competitive actions and dynamics in the digital age: An empirical investigation of social networking firms. *Information Systems Research*, 21, 3 (2010), 594-613.
40. Gómez-Solórzano, M., Tortoriello, M., and Soda, G. Instrumental and affective ties within the laboratory: The impact of informal cliques on innovative productivity. *Strategic Management Journal*, 40, 10 (2019), 1593-1609.
41. Huang, T., and Van Mieghem, J.A. Clickstream data and inventory management: Model and empirical analysis. *Production and Operations Management*, 23, 3 (2014), 333-347.
42. Hui, K.-L. Product variety under brand influence: An empirical investigation of personal computer demand. *Management Science*, 50, 5 (2004), 686-700.
43. Jabr, W., Ghoshal, A., Cheng, Y., and Pavlou, P. Maximizing Online Revisiting and Purchasing: A Clickstream-Based Approach to Enhancing Customer Lifetime Value. *Journal of Management Information Systems*, 40, 2 (2023), 470-502.

44. Johnson, E.J., Moe, W.W., Fader, P.S., Bellman, S., and Lohse, G.L. On the depth and dynamics of online search behavior. *Management science*, 50, 3 (2004), 299-308.
45. Kar, A.K., Angelopoulos, S., and Rao, H.R. Big data-driven theory building: Philosophies, guiding principles, and common traps. *International Journal of Information Management*, 71, 102661 (2023), 1-7.
46. Kim, J.B., Albuquerque, P., and Bronnenberg, B.J. Mapping online consumer search. *Journal of Marketing research*, 48, 1 (2011), 13-27.
47. Knoben, J., Oerlemans, L.A., Krijkamp, A.R., and Provan, K.G. What do they know? The antecedents of information accuracy differentials in interorganizational networks. *Organization Science*, 29, 3 (2018), 471-488.
48. Kossinets, G., and Watts, D.J. Empirical analysis of an evolving social network. *Science*, 311, 5757 (2006), 88-90.
49. Kossinets, G., and Watts, D.J. Origins of homophily in an evolving social network. *American journal of sociology*, 115, 2 (2009), 405-450.
50. Kushwaha, T., and Shankar, V. Are multichannel customers really more valuable? The moderating role of product category characteristics. *Journal of Marketing*, 77, 4 (2013), 67-85.
51. Lee, T.Y., and Bradlow, E.T. Automated marketing research using online customer reviews. *Journal of Marketing Research*, 48, 5 (2011), 881-894.
52. Li, B., Li, X., and Liu, H. Consumer preferences, cannibalization and competition: Evidence from the personal computer industry. *MIS Quarterly*, 42, 2 (2018), 661–678.
53. Li, H., Calder, C.A., and Cressie, N. Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model. *Geographical analysis*, 39, 4 (2007), 357-375.
54. Li, S., Liu, Y., and Feng, J. Who should own the data? The impact of data ownership shift from the service provider to consumers. *Journal of Management Information Systems*, 40, 2 (2021), 366-400.
55. Lin, C.F. Segmenting customer brand preference: demographic or psychographic. *Journal of Product & Brand Management* (2002).
56. Liu, X. Identification and efficient estimation of simultaneous equations network models. *Journal of Business & Economic Statistics*, 32, 4 (2014), 516-536.
57. Liu, Y., and Cui, T.H. The length of product line in distribution channels. *Marketing Science*, 29, 3 (2010), 474-482.
58. Liu, Y., Qian, Y., Jiang, Y., and Shang, J. Using favorite data to analyze asymmetric competition: Machine learning models. *European Journal of Operational Research*, 287, 2 (2020), 600-615.
59. Markóczy, L., Li Sun, S., Peng, M.W., Shi, W., and Ren, B. Social network contingency, symbolic management, and boundary stretching. *Strategic Management Journal*, 34, 11 (2013), 1367-1387.
60. Martínez-de-Albéniz, V., Planas, A., and Nasini, S. Using clickstream data to improve flash sales effectiveness. *Production and Operations Management*, 29, 11 (2020), 2508-2531.
61. McEvily, B., and Zaheer, A. Bridging ties: A source of firm heterogeneity in competitive capabilities. *Strategic management journal*, 20, 12 (1999), 1133-1156.
62. Mena, C., Karatzas, A., and Hansen, C. International trade resilience and the Covid-19 pandemic. *Journal of Business Research*, 138 (2022), 77-91.
63. Miranda, S., Berente, N., Seidel, S., Safadi, H., and Burton-Jones, A. Editor's Comments: Computationally Intensive Theory Construction: A Primer for Authors and Reviewers. *MIS Quarterly*, 46, 2 (2022), iii-xviii.
64. Moe, W.W., and Fader, P.S. Dynamic conversion behavior at e-commerce sites. *Management Science*, 50, 3 (2004), 326-335.
65. Moe, W.W. An empirical two-stage choice model with varying decision rules applied to internet clickstream data. *Journal of Marketing Research*, 43, 4 (2006), 680-692.
66. Moody, J., McFarland, D., and Bender-deMoll, S. Dynamic network visualization. *American journal of sociology*, 110, 4 (2005), 1206-1241.
67. Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M. Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31, 3 (2012), 521-543.
68. Newman, J.W., and Lockeman, B.D. Measuring prepurchase information seeking. *Journal of Consumer Research*, 2, 3 (1975), 216-222.
69. Opsahl, T., Agneessens, F., and Skvoretz, J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social networks*, 32, 3 (2010), 245-251.
70. Paulssen, M., and Bagozzi, R.P. Goal hierarchies as antecedents of market structure. *Psychology & Marketing*, 23, 8 (2006), 689-709.
71. Perera, S., and Tan, D. In search of the “Right Price” for air travel: First steps towards estimating granular price-demand elasticity. *Transportation Research Part A: Policy and Practice*, 130 (2019), 557-569.
72. Peter, J.P., Olson, J.C., and Grunert, K.G. Consumer behavior and marketing strategy. (1999).

73. Ravindran, K., Susarla, A., Mani, D., and Gurbaxani, V. Social capital and contract duration in buyer-supplier networks for information technology outsourcing. *Information Systems Research*, 26, 2 (2015), 379-397.
74. Reed, W.R. On the practice of lagging variables to avoid simultaneity. *Oxford Bulletin of Economics and Statistics*, 77, 6 (2015), 897-905.
75. Ringel, D.M., and Skiera, B. Visualizing asymmetric competition among more than 1,000 products using big search data. *Marketing Science*, 35, 3 (2016), 511-534.
76. Rowley, T., Behrens, D., and Krackhardt, D. Redundant governance structures: An analysis of structural and relational embeddedness in the steel and semiconductor industries. *Strategic management journal*, 21, 3 (2000), 369-386.
77. Sasidharan, S., Santhanam, R., Brass, D.J., and Sambamurthy, V. The effects of social network structure on enterprise systems success: A longitudinal multilevel analysis. *Information Systems Research*, 23, 3-part-1 (2012), 658-678.
78. Sismeiro, C., and Bucklin, R.E. Modeling purchase behavior at an e-commerce web site: A task-completion approach. *Journal of marketing research*, 41, 3 (2004), 306-323.
79. Sivakumar, K. Manifestations and measurement of asymmetric brand competition. *Journal of Business Research*, 57, 8 (2004), 813-820.
80. Stine, R.A. Graphical interpretation of variance inflation factors. *The American Statistician*, 49, 1 (1995), 53-56.
81. Stokburger-Sauer, N.E., and Teichmann, K. Is luxury just a female thing? The role of gender in luxury brand consumption. *Journal of business research*, 66, 7 (2013), 889-896.
82. Struijk, M., Angelopoulos, S., Ou, C., and Davison, R.M. Influencing information quality: Evidence from a military organization. *European Conference on Information Systems*: Association for Information Systems, 2020.
83. Struijk, M., Ou, C.X., Davison, R.M., and Angelopoulos, S. Putting the IS back into IS research. *Information Systems Journal*, 32, 3 (2022), 1-4.
84. Struijk, M., Angelopoulos, S., Ou, C.X., and Davison, R.M. Navigating Digital Transformation Through an Information Quality Strategy: Evidence From a Military Organization. *Information Systems Journal*, 33, 4 (2023).
85. Sun, C., Zhang, X., Zhou, Y.-W., and Cao, B. Pricing, financing and channel structure for capital-constrained dual-channel supply chains with product heterogeneity. *International Journal of Production Economics*, 253 (2022), 108591.
86. Tan, T.F., Netessine, S., and Hitt, L. Is Tom Cruise threatened? An empirical study of the impact of product variety on demand concentration. *Information Systems Research*, 28, 3 (2017), 643-660.
87. Tan, X., Jin, F., and Dennis, A.R. How Appreciation and Attention Affect Contributions to Electronic Networks of Practice. *Journal of Management Information Systems*, 39, 4 (2022), 1037-1063.
88. Tortoriello, M. The social underpinnings of absorptive capacity: The moderating effects of structural holes on innovation generation based on external knowledge. *Strategic Management Journal*, 36, 4 (2015), 586-597.
89. Upson, J.W., Ketchen Jr, D.J., Connelly, B.L., and Ranft, A.L. Competitor analysis and foothold moves. *Academy of Management Journal*, 55, 1 (2012), 93-110.
90. Van Mierlo, T. The 1% rule in four digital health social networks: an observational study. *Journal of medical Internet research*, 16, 2 (2014), e2966.
91. Ver Hoef, J.M., Peterson, E.E., Hooten, M.B., Hanks, E.M., and Fortin, M.J. Spatial autoregressive models for statistical inference from ecological data. *Ecological Monographs*, 88, 1 (2018), 36-59.
92. Wasserman, S., and Faust, K. Social network analysis: Methods and applications. (1994).
93. Wu, L. Social network effects on productivity and job security: Evidence from the adoption of a social networking tool. *Information systems research*, 24, 1 (2013), 30-51.
94. Yang, S., and Allenby, G.M. Modeling interdependent consumer preferences. *Journal of Marketing Research*, 40, 3 (2003), 282-294.
95. Yang, Y., and Fik, T. Spatial effects in regional tourism growth. *Annals of Tourism Research*, 46 (2014), 144-162.
96. Zhang, J.J., Fang, X., and Liu Sheng, O.R. Online consumer search depth: Theories and new findings. *Journal of Management Information Systems*, 23, 3 (2006), 71-95.
97. Zheng, Z., Fader, P., and Padmanabhan, B. From business intelligence to competitive intelligence: Inferring competitive measures using augmented site-centric data. *Information Systems Research*, 23, 3-part-1 (2012), 698-720.
98. Zuo, M., Angelopoulos, S., Liang, Z., and Ou, C.X. Blazing the Trail: Considering Browsing Path Dependence in Online Service Response Strategy. *Information Systems Frontiers* (2022), 1-15.

APPENDIX I

To clearly demonstrate the connections and differences between this study and the prior ones on intra- and inter-brand competition, we conduct comparative analysis from six aspects, namely objective, source, type of data, competition and normalization dynamics, reasons for dynamic changes, and large product categories. The results of our analysis are shown in Table I below. According to this table, our study considers the *dynamic* asymmetric competition in various time-windows from the perspective of multiple product categories, and further analyzes the network-related factors that can influence the degree of asymmetric competition in the context of online marketplaces.

Table V: *Overview of Previous Studies and Comparison with Our Study*

Study	Netzer et al. [67]	Ringel and Skiera [75]	Li et al. [52]	Our study
Objective	Mine user generated content to identify market structure and competitive landscape	Use search data to analyze asymmetric competition	Identify cannibalization and competition effects via consumers' search behaviors and brand preferences	Identify cannibalization and competition effects through consumers' search behavior, and analyze the reasons for changes in the degree of competition
Source	Discussion forums	Product- and price comparison	Online and offline retail information	Sequential clicking and purchasing behavior
Type of data	User Generated Content	Consumer consideration sets	Sales and survey data	Consumer consideration set
Competition and cannibalization dynamics	No	No	No	Yes
Reasons for dynamic changes	No	No	No	Yes
Large product categories	No	Yes	No	Yes

APPENDIX II

To measure the duration of consumers' search from the beginning of the process to their final purchase, we conducted a preliminary statistical analysis on the dataset of our study. In doing so, we take the duration between two purchases from consumers as a sample, and the statistical results are shown in Figure I. The x-axis represents the average duration between any two purchases by all consumers, which represents the average value searched by consumers during a purchase process, where the *y*-axis represents the proportion of consumers corresponding to the search duration. The Figure I demonstrates that 85% of consumers searches have duration of 14 days before proceeding to the final purchase.

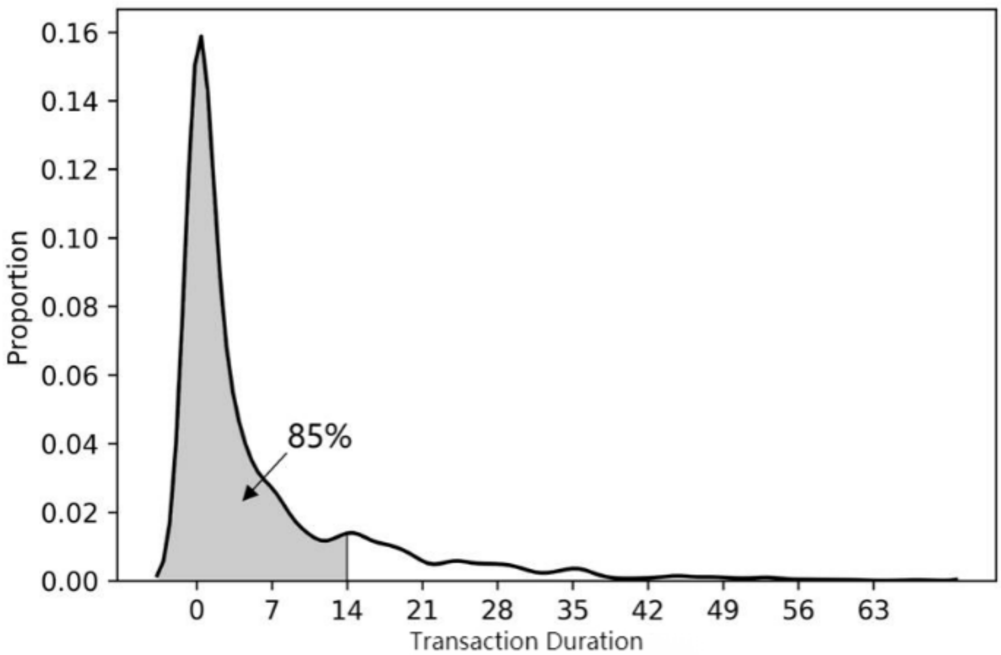

Figure V: *Probability Distribution of the Search Duration in a Purchase*

APPENDIX III

By visualizing the undirected competition network, we generate the network graph presented in Figure II. The graph consists of 52 nodes, representing the brands, and 916 ties, which represent competing dyads. The network density is 0.69, indicating that approximately 69% of the brands have a competitive relationship. The average degree of the undirected competition network is 35.23, and the average path length is 1.35. These metrics suggest that brands within the same product category are closely interconnected, highlighting the intense competition among the brands in our study.

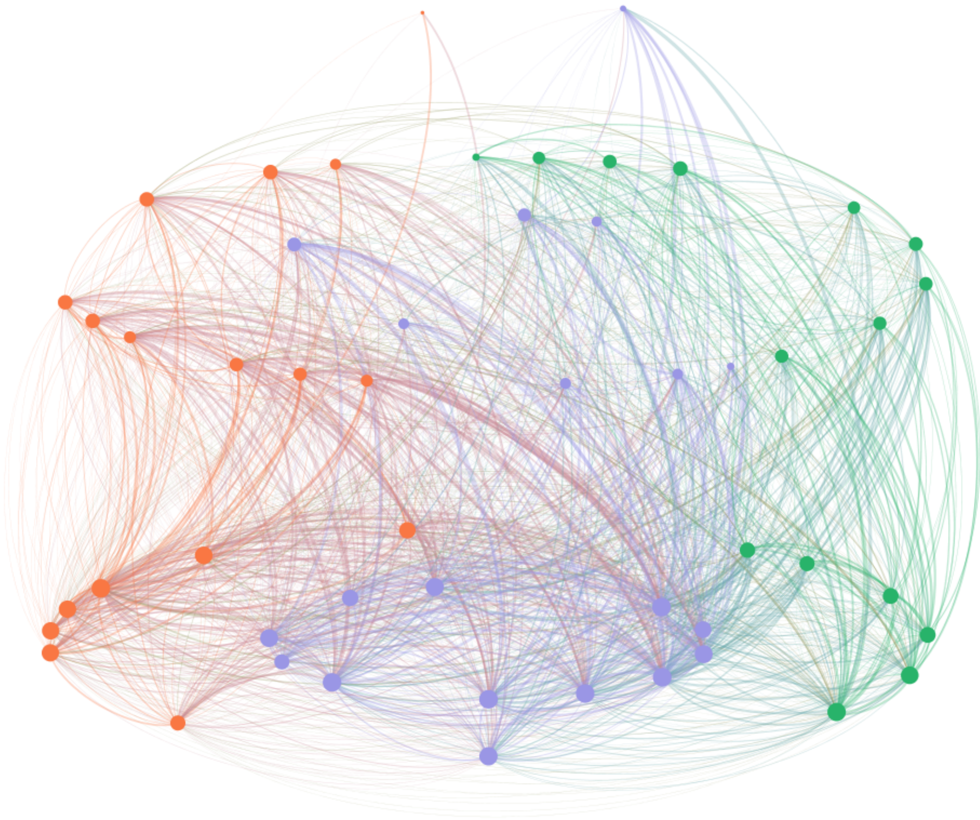

Figure II: *Diagram of Competition between Any Two Brands*

Note: The nodes represent 52 brands, and their size indicates sales volume. The ties indicate a competitive relationship between brands; the color of the ties represents the market segments.

To cluster the competition relations among brands and products, we utilize the Louvain community detection approach with resolution parameters. Additionally, we employ in-degree centrality to measure the degree of competition within the first time-window. The clustering results are presented in Figure III, where each node represents a brand, and its size corresponds to its sales volume. Market segments are indicated by different colors, while ties between nodes indicate that brands are included in a consumer's consideration set within the given time-window. Although we can create nine such network graphs, one for each time-window in our study, for brevity, we present the results only for the first time-window.

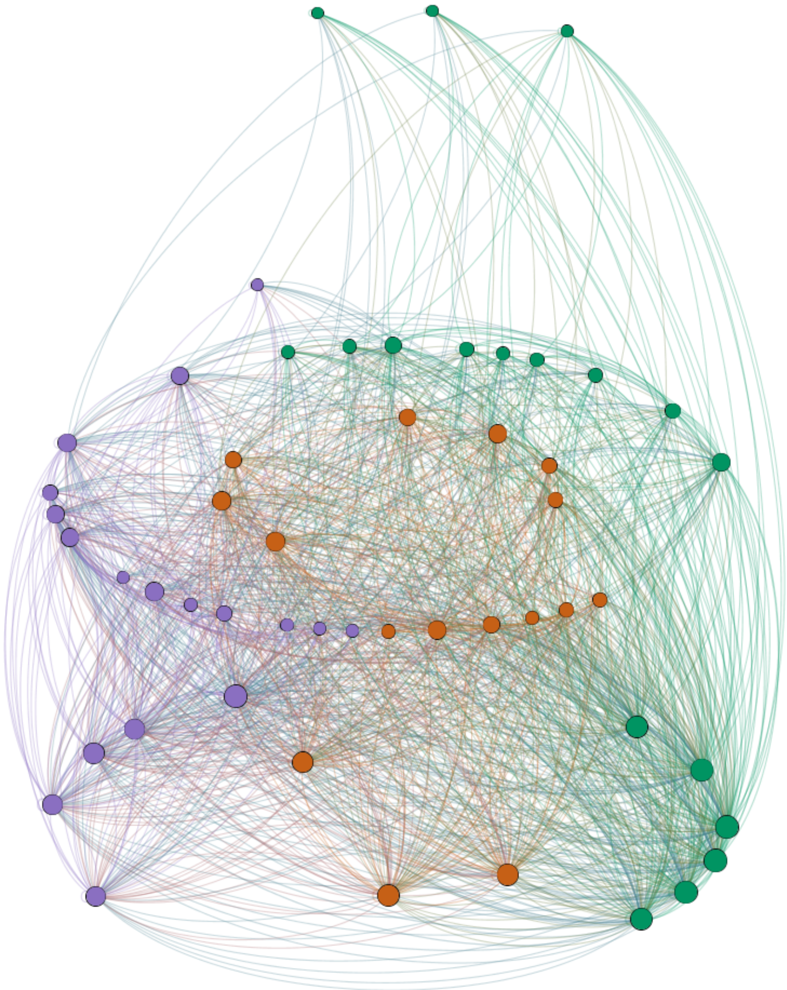

Figure III: *Asymmetric Competition between Brands in the First Time-window*

Note: Each node represents a brand, and their size indicates sales volume. The ties indicate a competitive relationship between brands; the color of the ties represents the market segments.

APPENDIX IV

To understand the purchases situation of products in the same category from various brands, we conducted purchases statistics on 52 brands, and the results are shown in Figure IV. The x-axis represents the ranking of brand purchases volume, while the y-axis represents the brand's purchases volume. We conducted statistical analysis on the purchase situation of the top 18 brands, and the results show that their sales accounted for 90% of the overall market share.

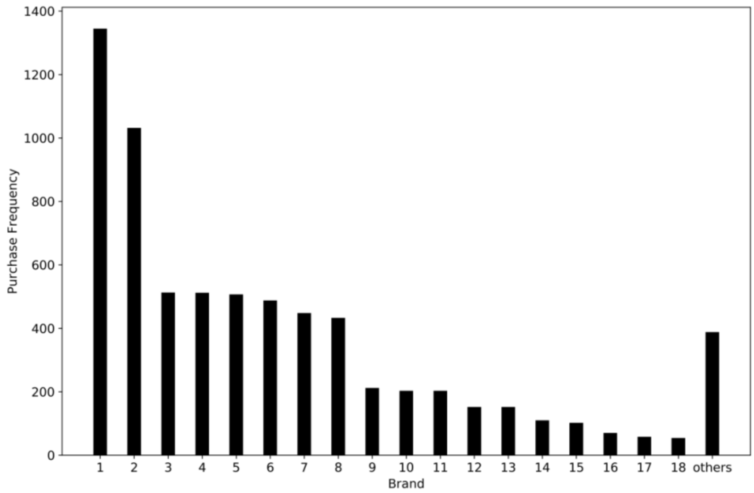

Figure IV: *The Distribution of Brand-purchases*

Note: The sample distribution of brands with purchase records shows a clear right tail. The small spike at the right tail is due to the truncation of the right tail above the 90th percentile.

APPENDIX V

The degree of competition for the top 18 brands is shown in Figure V. The x-axis represents the time-windows, the y-axis represents the degree of competition, and each brand is represented by a distinct color. To measure the degree of competition, we use in-degree centrality in the inter-brand network. The degree of competition exhibits fluctuations within the time-windows, which can be categorized into two phases. Most brands experience fluctuations within a range of five points of scale, while a few brands exhibit larger fluctuations, as indicated by the dashed box. For instance, consider brand '165': in time-window 3, its degree of competition approaches 40, whereas in time-window 5, it drops below 5. Significantly fluctuating degree of competition can pose operational risks for brands.

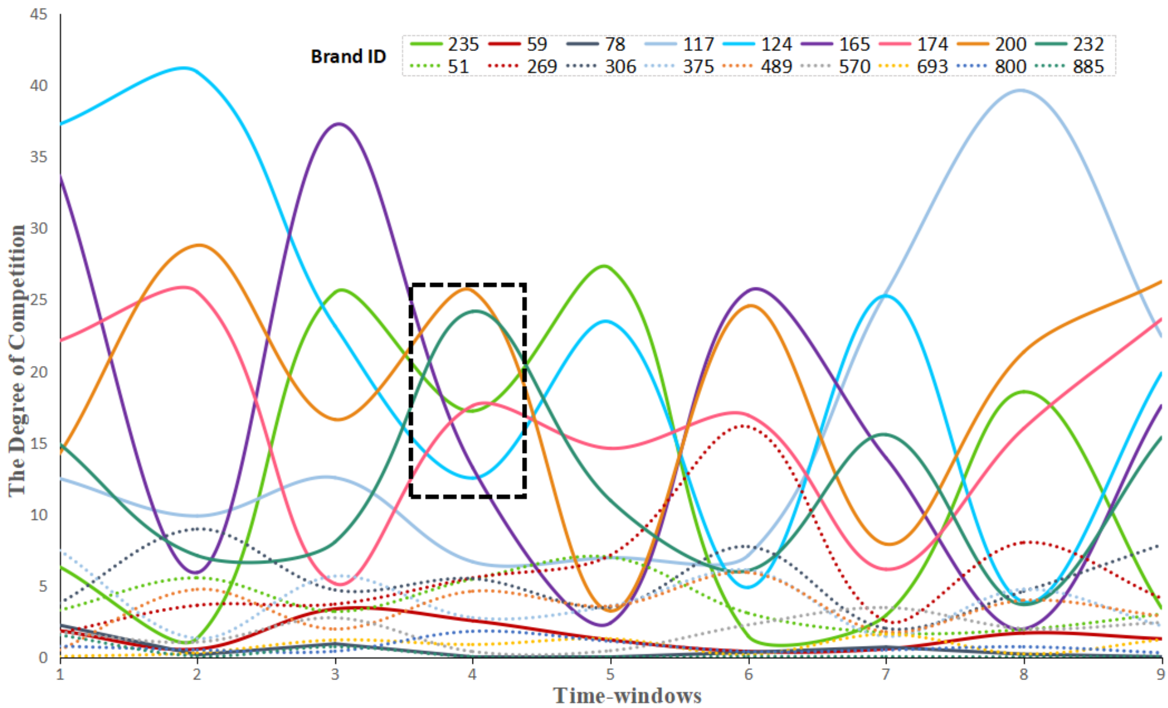

Figure V: *The Degree of Competition of Each Brand in the Period of Nine Time-windows*

Note: The different colors represent different brands as indicated in the upper right corner.

APPENDIX VI

We select the products with the top 90% sales volume from each brand and analyze the intra- and inter-brand competition in different time-windows. The outcome is depicted in Figure VI, which consists of five directed network graphs representing intra- and inter-brand competition. In these graphs, each node represents a product. Nodes belonging to the same brand have the same color, as illustrated below Figure VIa. Bilateral ties indicate the presence of intra- and inter-brand competition, while unilateral ties signify that a specific product poses a competitive threat to others without reciprocation. The weight of the ties indicates the extent of competition. The number and weight of ties between products within a brand surpass those between brands, revealing a higher level of intra-brand competition compared to inter-brand one, and suggesting that consumers tend to consider products within the same brand rather than across different brands. By examining the network in different time-windows it becomes apparent that the ties and their thickness exhibit dynamic changes, demonstrating that the degree of intra- and inter-brand competition undergoes continuous fluctuations.

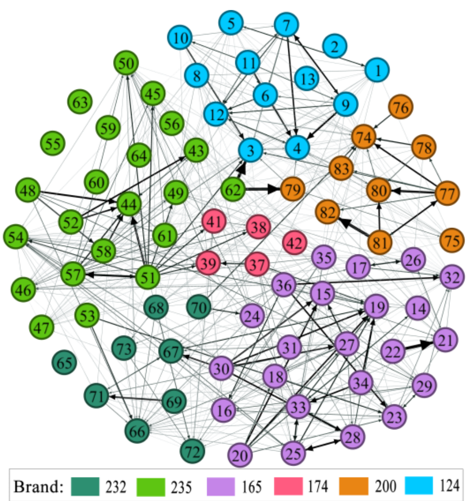

Figure VIa First Time-window

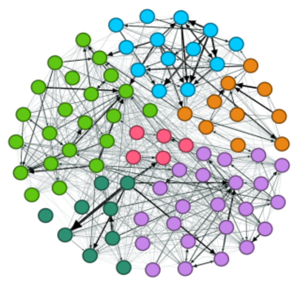

Figure VIb Third Time-window

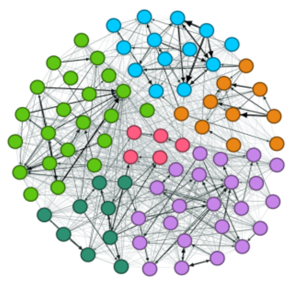

Figure VIc Fifth Time-window

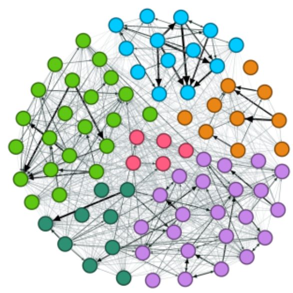

Figure VId Seventh Time-window

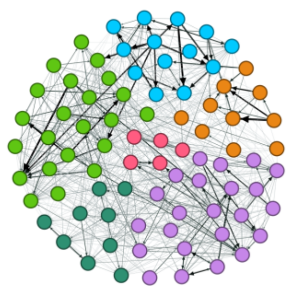

Figure VIe Ninth Time-window

Figure VI: *Intra- and Inter-brand competition in different time-windows*

APPENDIX VII

To verify that intra-band competition is much higher than inter-band one, we divided the 52 brands into three categories based on their sales volume (high, medium, low), and randomly selected one brand from each category. For each category of products, we build intra-band competition network with different time windows, calculate the indicators of each product, and analyze the relation between the indicators and the sales volume with a mixed regression. The results are shown in Table II, where the value of R square is 0.839, indicating that 83.9% of sales fluctuations are contributed by intra-brand competition.

Table II: *The Impact of Intra-Brand Relations on Brand Sales*

	Regression coefficients	**VIF**	**R square**
Constant	19.886** (6.526)		
1-Constrain	-105.687*** (11.888)	3.478	
Betweenness Centrality	-22.262 (28.356)	1.175	0.839
Degree Centrality	2.934*** (0.083)	3.719	

APPENDIX VIII

To test the correlation between all variables of interest, we conducted correlation tests on the DV, IV, and CV, and the results are shown in Table III. The following table shows that the maximum correlation coefficient is 0.66, indicating a correlation of 0.66 between Degree Centrality and 1-Constraint. The minimum correlation coefficient is 0.02, indicating a correlation of 0.02 between age and log sales, while the correlation between other variables is between 0.66 and 0.02.

Table III: *Correlation Matrix*

		(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Log Sales	(1)	1.00						
1-Constrain	(2)	-0.04	1.00					
Degree Centrality	(3)	0.37	0.66	1.00				
Betweenness Centrality	(4)	0.42	0.50	0.49	1.00			
Gender	(5)	0.18	-0.08	-0.02	-0.01	1.00		
Age	(6)	-0.02	-0.13	-0.12	-0.11	-0.16	1.00	
Membership level	(7)	0.42	0.10	0.19	0.28	0.16	-0.16	1.00

APPENDIX IX

To test multicollinearity among IV and CV, we performed a variance inflation factor (VIF) test, and the results are shown in Table IV. The following table shows that the maximum value of VIF for all variable is 2.01, which is far less than its threshold of 10, indicating that multicollinearity among the IV and CV is not an issue.

Table IV: *Variance Inflation Factor (VIF) of Each Variable*

	Variables	VIF	1/VIF
IV	1-Constrain	2.01	0.49
	Degree Centrality	1.95	0.51
	Betweenness Centrality	1.52	0.66
CV	Gender	1.15	0.87
	Age	1.07	0.93
	Membership level	1.06	0.94
Mean VIF		1.46	

Note: Since the maximum value of VIF in Table IV is 2.01, which is far less than 10, multicollinearity is not an issue in this study.

APPENDIX X

Model Specification

Spatial Auto-Regressive Models

There are three types of spatial panel data models that consider state dependence (Yang and Fik 2014):

1. Spatial auto-regressive (SAR) model, which considers spatial dependence on the DV as shown in (8):

$$SAR: Y_{jt} = \lambda \sum_{k=1}^{N} W_{jk} Y_{kt} + X_{jt-1}\beta + \mu_j + \delta_t + \epsilon_{jt}, (8)$$

2. Spatial Error Model (SEM), which considers spatial dependence on error terms as shown in (9):

$$SEM: Y_{jt} = X_{jt-1}\beta + \mu_j + \delta_t + \varphi_{jt}$$

$$where \varphi_{jt} = n \sum_{k=1}^{N} W_{jk} \varphi_{jt} + \epsilon_{jt}, (9)$$

3. Space Durbin Model (SDM), which considers spatial dependence through the lagged DV and the lagged IV as shown in (10):

$$SDM: Y_{jt} = \lambda \sum_{k=1}^{N} W_{jk} Y_{kt} + X_{jt-1}\beta + \sum_{k=1}^{N} W_{jk} X_{kt-1}\vartheta + \mu_j + \delta_t + \epsilon_{jt}, (10)$$

Model Selection

Before analyzing the models, the first step is to determine whether to use the spatial time fixed effect model or the spatial time random effect model. We use Hausman specification test for detection, in which if the Hausmann statistic is negative, the spatial time random effect model is used; otherwise, the spatial time fixed effect model is used (Wagner 2005; Su and Lin 2014). The statistical results are shown in Table V. The parameter value of chi2 is -10.48, indicating the Hausmann statistic is negative. As a result, the random effects model should be used for parameter estimation.

Table V: *Hausman Specification Test Results*

Variable		(b) SDM-fe	(B) SDM-re	(b-B) Difference
IV	1-Constrain	-1.912	-4.236	2.323
	Degree Centrality	0.010	0.018	-0.008
	Betweenness Centrality	0.001	0.002	-0.001
DV	Gender	0.395	0.408	-0.013
	Age	0.069	0.191	-0.121
	Membership level	0.079	0.108	-0.029

b=consistent under Ho and Ha; B=inconsistent under Ha, efficient under Ho;
Test: Ho: difference in coefficients not systematic
Chi2(6) = -10.48

Since the SAR random model is in the form of SDM, SEM, and SAR, different parameter estimations are obtained based on different models as shown in Model 3, 4, and 5 in Table VI. The SAR model is a simplified case of SDM as it excludes lag IV and only includes the lag term of the DV. While the SEM is also a simplification of the SDM, the spatial dependence of the IV and DV in the SDM is simplified by the SAR coefficients in the SEM. Therefore, we use Wald statistic (Yang and Fik 2014) to test whether the SDM can be simplified to the SAR or SEM. The Wald test results of the SDM and SEM are represented by Wald1, while the ones of the SDM and SAR are represented by Wald2. As shown in Table VI, the values of Wald statistic are not significant. As a result, the SDM can be simplified to SEM and SAR by accepting the null hypothesis. Table VI demonstrates that the parameter estimation results of the SEM and the SAR model are not very different, but the SAR model has a lower AIC value and a higher R-squared than that of SEM, indicating that the SAR model has a better goodness-of-fit for the log of sales. Thus, we use the SAR model directly in our analysis.

Table VI: *Estimation Results for OLS and SAR*

Variables		Model 1 OLS	Model 2 OLS	Model 3 SDM	Model 4 SEM	Model 5 SAR
CV	**Gender**	0.998*** (0.323)	0.879*** (0.246)	0.408** (0.176)	0.396** (0.157)	0.393** (0.160)
	Membership level	3.259*** (0.331)	2.040*** (0.262)	0.190 (0.181)	0.190 (0.177)	0.185 (0.175)
	Age	0.308 (0.201)	0.323** (0.153)	0.108 (0.133)	0.078 (0.189)	0.083 (0.184)
IV	**1-Constrain**		-53.905*** (3.788)	-4.325 (2.896)	-4.752* (2.541)	-4.244* (2.337)
	Degree Centrality		0.157*** (0.012)	0.018* (0.009)	0.020*** (0.007)	0.017** (0.007)
	Betweenness Centrality		0.022*** (0.002)	0.002 (0.001)	0.002*** (0.001)	0.002** (0.001)
Network	**Rho**			0.574*** (0.106)	0.540*** (0.070)	0.589*** (0.063)
	Constant	0.850*** (0.115)	38.099*** (3.032)	13.866 (12.786)	4.180** (1.816)	3.306* (1.721)
R square		**0.200**	**0.539**	**0.594**	**0.673**	**0.687**
Wald1					**3.820**	
Wald2						**3.810**
AIC				**800.048**	**797.648**	**791.847**

*Note: Rho represents the correlation of a competitive relation between a neighbor brand and a focal brand. R square represents the fitting effect of the corresponding model on actual sales. Wald1 refers to the Wald test between SDM and SEM models, while Wald2 refers to the Wald test between SDE and SAR models. AIC represents the goodness-of-fit of the spatial auto-regressive model, and the lower the value the better the goodness-of-fit. Standard errors are listed in parentheses; Path coefficients are listed outside the parentheses. * p<0.1; ** p<0.05; *** p<0.01*

APPENDIX XI

Robustness Check

We use consumer clicking and purchasing behavior data of different product categories to verify whether the three IV affect the brand sales in the same way as the original dataset. The results are shown in Model 2 in Table VII.

Table VII: *Robustness Check (Normalized Variables)*

Variables		Model 1 SAR (Original data set)	Model 2 SAR (Test data set)
CV	**Gender**	0.052** (0.021)	- 0.058 (0.047)
	Membership level	0.024 (0.023)	0.067 (0.091)
	Age	0.018 (0.039)	0.079 (0.075)
IV	**1-Constrain**	-0.050* (0.027)	**- 0.236***** (0.058)
	Degree Centrality	0.064** (0.026)	**0.347***** (0.058)
	Betweenness Centrality	0.037** (0.014)	0.062* (0.036)
Network	**Rho**	**0.589***** (0.064)	**0.448***** (0.072)
	Constant	0.187 (0.122)	**-0.308***** (0.115)
R square		0.688	0.598
AIC		420.515	539.747

*Note: Standard errors are listed in parentheses. Path coefficients are listed outside the parentheses. R square represents the fitting effect of the corresponding model on actual sales. AIC represents the goodness-of-fit of the SAR model. The lower the AIC value the better the goodness-of-fit. * p<0.1; ** p<0.05; *** p<0.01.*

Analysis Results of Different Duration of the Sample Period

We conduct further robustness checks by dividing weeks into weekdays/weekends. Since there are ten weeks in total, but there are only weekdays in the last week, the 10-week data can be divided into 21 time-windows. We selected 53 brands that are clicked every weekday and weekend, where 2,517,090 clicks were made by 53,482 consumers, and the number of click products was 5,292,436. We form 21 networks with asymmetric competition among brands. The IV and CV are statistically calculated, and the correlation coefficients are shown in Table VIII. Again, there are differences in the selection results of SAR models due to different data structures. We use Wald test (Yang and Fik 2014) to check if the SDM can be simplified to SAR or SEM. The Wald test results of the SDM and SAR are represented by Wald1, while for SDM and SEM are represented by Wald2. According to the Wald test results in Table IX, the values of Wald statistic are significant. The SDM thus cannot be simplified to SEM and SAR by rejecting the null hypothesis.

Table VIII: *Correlation Table for All Possible Variables*

		(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Weekend	(1)	1.00									
Gender	(2)	-0.05	1.00								
Age	(3)	0.01	0.02	1.00							
Membership level	(4)	0.06	0.20	-0.20	1.00						
1-Constrain	(5)	-0.09	-0.04	-0.10	0.14	1.00					
Betweenness Centrality	(6)	-0.10	0.03	-0.07	0.11	-0.01	1.00				
In-degree Centrality	(7)	-0.18	-0.03	-0.09	0.11	0.48	0.19	1.00			
Out-degree Centrality	(8)	-0.23	-0.09	-0.02	0.04	0.50	0.16	0.81	1.00		
Closeness Centrality	(9)	-0.19	-0.01	-0.14	0.15	0.33	0.45	0.82	0.59	1.00	
Eigenvector Centrality	(10)	-0.19	-0.01	-0.10	0.14	0.46	0.24	0.94	0.73	0.80	1.00

Table IX: *The Results of Wald Test*

	Wald1	Wald2	P-value
SDM vs SAR	43.59		0.000
SDM vs SEM		27.81	0.0005

According to the above table of correlation coefficients among all variables, we first exclude the eigenvector centrality indicator. Only one of the in-degree and out-degree centrality indicators can be selected. The statistical results of the SDM model are shown in Table X:

Table X: *Results of the SDM Model*

	Variables	Model 1	Model 2
		SDM	SDM
CV	**Weekend**	0.31 (0.38)	0.18 (0.39)
	Gender	-1.49 (0.94)	-1.56 (0.95)
	Age	0.47 (0.90)	0.92 (0.82)
	Membership level	-0.11 (0.91)	0.31 (0.89)
IV	**1-Constrain**	-0.93** (0.33)	-0.65** (0.31)
	Betweenness Centrality	-0.64 (5.55)	-16.64** (6.34)
	In-degree Centrality	6.88*** (1.88)	
	Out-degree Centrality		-1.93 (1.29)
	Closeness Centrality		0.23*** (0.05)
	Constant	-2.26 (15.33)	-0.74 (14.75)
R square		**0.81**	**0.80**

*Note: Standard errors are listed in parentheses. Path coefficients are listed outside the parentheses. R square represents the fitting effect of the corresponding model on actual sales. AIC represents the goodness-of-fit of the SAR model. The lower the AIC value the better the goodness-of-fit. * p<0.1; ** p<0.05; *** p<0.01.*

The IV in Table X is the network indicators of each period. The DV is the sales volume of each brand in the same period. In line with Ravindran et al. (2015), since CV is not significant, we directly used the SAR model for two attempts. First, we used the SAR model to recalculate Model 1 in Table VII. The results are in Model 1 in Table XI. Second, we used the SAR model to predict sales volume for the next time-window. The results are shown in model 2 in Table XI.

Table XI: *Results of the SAR Model*

Variables		**Model 1**	**Model 2**
		SAR	**SAR**
CV	**Weekend**	0.30 (0.38)	4.33*** (0.68)
	Gender	-1.52* (0.88)	-0.65 (1.51)
	Age	0.70 (0.88)	-1.08 (1.16)
	Membership level	-0.32 (0.94)	0.28 (1.31)
IV	**1-Constrain**	-3.74** (1.47)	0.51 (0.74)
	Betweenness Centrality	-2.54 (5.52)	-0.07 (6.22)
	In-degree Centrality	8.68*** (1.73)	2.79** (1.37)
	Constant	-5.73** (1.88)	-6.86 (11.65)
R square		**0.80**	**0.72**

*Note: Standard errors are listed in parentheses. Path coefficients are listed outside the parentheses. R square represents the fitting effect of the corresponding model on actual sales. AIC represents the goodness-of-fit of the SAR model. The lower the AIC value the better the goodness-of-fit. * $p<0.1$; ** $p<0.05$; *** $p<0.01$.*