Journal of
Theoretical and Applied
Electronic Commerce Research

Article

The Impact of Topological Structure, Product Category, and Online Reviews on Co-Purchase: A Network Perspective

Hongming Gao

School of Management, Guangzhou University, Guangzhou 510006, China; hmgao@gzhu.edu.cn

Abstract: Understanding the relationships within product co-purchasing is crucial for designing effective cross-selling and recommendation systems in e-commerce. While researchers often detect co-purchase rules based on product attributes, this study explores the influence of consumer behavior preferences and electronic word-of-mouth (eWOM) on co-purchase formation by analyzing the topological network structure of products. Data were collected from a major Chinese e-retailer and analyzed using an exponential random graph model (ERGM) to identify the factors affecting the formation of follow-up purchases between products: the role of topological structure, product category, and online product reviews. The results showed that the co-purchase network was a sparse small-world network, with a product degree of centrality that positively impacted its sales volume within the network, suggesting a concentration effect. Cross-category purchases significantly contribute to the formation of co-purchase relationships, with a differential homophily effect. Positive ratings and review volumes were found to be key factors impacting this co-purchase formation. In addition, a higher inconsistency of positive ratings among products decreases the likelihood of co-purchase. These findings contribute to the literature on eWOM and electronic networks, and have valuable implications for e-commerce managers.

Keywords: co-purchase network; eWOM; topological structure; product category; online reviews; ERGM

Citation: Gao, H. The Impact of Topological Structure, Product Category, and Online Reviews on Co-Purchase: A Network Perspective. *J. Theor. Appl. Electron. Commer. Res.* **2023**, *18*, 548–570. https://doi.org/10.3390/jtaer18010028

Academic Editors: Mirjana Pejić-Bach and María Teresa Ballestar

Received: 6 February 2023
Revised: 24 February 2023
Accepted: 9 March 2023
Published: 10 March 2023

1. Introduction

Marketing applications that utilize electronic networks are becoming more prevalent within various industries. They help to detect users' preferences and improve business performance through social network analysis. As outcomes of homophily [1,2], the connected people disclose their demand dependency on items and are more inclined to respond similarly. Specifically, managers often implement network-based marketing tactics by using recommendations or product hyperlinks for individuals, in order to reveal collective preferences and improve their targeting choices [3]. Examples of this include displaying "also-bought" hyperlinks in online retailers, "also-liked" hyperlinks in music suppliers, and "also-viewed" hyperlinks in video streaming platforms and news consumption areas.

The most prevalent network application in e-commerce is co-purchase hyperlink design. Products are represented as nodes and co-purchase hyperlinks are referred to as recommendation lists under the label "customers who bought this also bought . . . " [4–7]. As an example, when viewing a Huawei P40 mobile phone (the focal product node), its co-purchase hyperlinks are displayed on both Amazon (Figure 1) and JD.com (Figure 2). The focal product and one of the products in its co-purchase hyperlink form a dyadic product-to-product level. Previous studies have examined the impact of demand spillovers from these co-purchase-network-based recommendation systems [8–10]. The evidence shows that co-purchase-network-based recommendations lead to significant sales increases between the focal product and its co-purchase hyperlinked products [11]. Furthermore, pattern mining for interest [12,13] and inter-competition among the products [14–17] in the co-purchase network have been researched. However, these systems can also restrict users in a "filter bubble" and negatively impact sales diversity. Optimizing product allocation and bundle arrangements within recommendation hyperlinks, in order to improve sales and cross-selling, is of great importance to both complex network research and e-commerce practices.

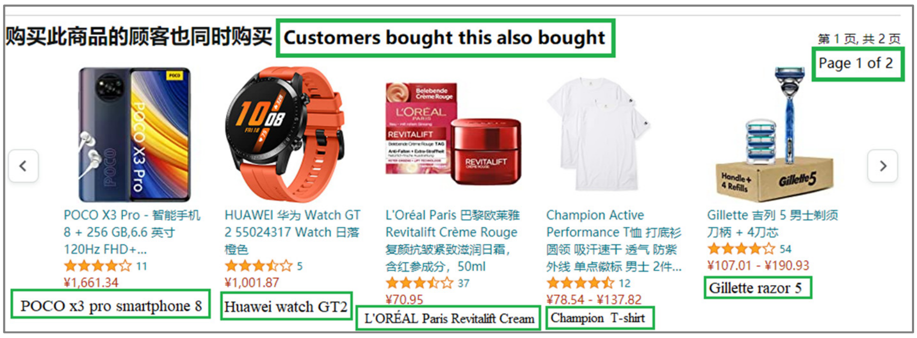


Figure 1. Co-purchase hyperlinks on Amazon (Amazon, available at: https://www.amazon.cn/dp/B085YSJ5YH/ref=pd_rhf_gw_p_img_1?_encoding=UTF8&psc=1&refRID=E4D3BKMA1EHPPZ9QVWX5, assessed on 6 August 2021).

Figure 2. Co-purchase hyperlinks on JD.com (JD.com, available at: https://cart.jd.com/addToCart.html?rcd=1&pid =10027944186268&pc=1&eb=1&rid=1628245641807&em= assessed on 6 August 2021).

The discovery of co-purchase relations is the basis for optimizing recommended hyperlink design. Prior research has primarily focused on the post hoc impact of co-purchase-network-based recommendation systems [4–7,9,10]. There are only a limited number of studies that examine the evolution of co-purchase relations. These studies have analyzed internal network factors, such as the endogenous attributes of the networks and the exogenous attributes of the products/nodes [12,14,17]. eWOM, a form of external influence on the co-purchase network, refers to reviews and opinions from a global audience [18,19]. Researchers often study the impact of co-purchase-network-based recommendations on customer satisfaction, measured through online product reviews. A complementary recommender is likely to result in diverging product ratings [20]. Furthermore, the ratings of a focal product are influenced by its network neighbors in the co-purchase network [21]. In turn, the convergence of eWOM ratings leads to an increase in the total sales for both products [22]. However, understanding of the directional impact of eWOM on the formation of co-purchase relations is still scarce in recent studies [8,20,21]. Managing eWOM in the co-purchase network is a relatively new topic. To the best of our understanding, this research is the inaugural exploration of the impact that eWOM valence and volume in online reviews have on the formation of product-to-product co-purchase relations.

Considering the above research gaps, an explanatory framework for co-purchase network formation is therefore of utmost importance. Hence, this study constructs a co-purchase network by linking products by the same individual users from a real-world dataset. Afterward, we examine the network topological structure by identifying the consumer buying patterns. In addition, we also adopt an ERGM approach by discovering the impacts of the topological structure, product category, and online reviews on co-purchase formation, which prioritizes statistical inference and the explanation of network formation [23,24]. Specifically, the research questions raised in this study are as follows:

RQ1: What are the topological properties of the co-purchase network?

RQ2: Does cross-product category purchase impact the formation of co-purchasing? Through what combination path?

RQ3: What is the impact of online reviews on the formation of co-purchasing? What is the directionality of the impact?

The understanding of the formation of co-purchase relationships at the product-to-product level is vital for advancing the research on electronic networks. Our contribution to the literature on electronic networks is threefold. Firstly, we extend the discourse on demand dependency by presenting an evolutionary framework to comprehend co-purchase decisions at the product-to-product level. Secondly, we expand the existing knowledge on the statistical mechanics of consumer buying patterns by exploring the characteristics of product cross-category purchase behaviors. Lastly, we include the review volume and positive ratings to obtain novel measures of eWOM, and examine the impact of eWOM consistency from external audiences on the co-purchase relationships between products. This study also provides valuable insights for e-retailers to enhance their explainable cross-selling and recommendation hyperlink systems.

The structure of the study is as follows. In Section 2, we review the relevant literature and provide a discussion. The proposed research framework and approach are outlined in Section 3. In Section 4, we conduct a series of experiments and present the results. Lastly, in Sections 5 and 6, we offer a discussion, a conclusion, and identify the potential avenues for future work.

2. Related Work

2.1. Co-Purchase Network Analysis

The topological structure of the product co-purchase network has been exploited to investigate the economic effects of its recommendation systems. Kumar and Hosanagar [8] found that the average daily number of page views increased due to the co-purchase recommendation network, incurring a higher user engagement and product exposure during online shopping. The outgoing hyperlinks from a product to its complements or substitutes create demand spillovers, leading to a significant increase in total sales [9,10]. On the other hand, there is a trade-off between sales increase and sales diversity [4,25]. Collaborative filters, a type of co-purchase recommendation system [26], may harm sales diversity [4–7]. Because such systems confine users in a “filter bubble”, where popular items are reinforced instead of those with limited purchase records [4,6]. Since we study e-commerce platforms, we use the terms user and consumer interchange.

Another area of research on co-purchase networks is in finding product relationships and positioning virtual shelves according to perceptual consumer preferences [11]. Pattern mining for interest [12,13] and the inter-competition among products [14–17] are also representatives in this research stream. Prior studies have focused on identifying the right products at the right positions between the focal products and their related ones for cross-selling or boosting sales. For example, Kim et al. analyzed the topological characteristics and structure of co-purchase networks and found that products with a higher degree of centrality are effective in promoting sales or cross-selling [12]. Researchers also mined a set of influential products within the network to increase the influence spread and lift the sales of other products [14]. SNA-based methods are proposed to conduct market basket analysis and identify two characteristics: the interrelationships among products that help to foster cross-selling and not just complementary or substitutive bundling, and the role of product communities that optimizes the product position allocation and collocations based on the preferences of different consumer groups [17,27].

Despite the research on post hoc effects (such as sales volume and diversity), there is limited understanding of the formation process of the co-purchase network, specifically its topological structures, cross-category product purchase patterns, and the formation of co-purchase relations between a focal product and another subsequent product.

2.2. Co-Purchase and eWOM

Previous studies have examined the internal factors affecting networks, such as the intrinsic attributes of the network (such as community clustering, coefficient, and degree distribution) and the extrinsic attributes of the products/nodes (such as price and purchasing frequency) [12,14,17]. With regard to the co-purchase network, one external factor is eWOM, which refers to reviews and opinions from a global audience on e-commerce platforms [18,19,28].

To further understand the relationship between online product reviews and co-purchase networks, researchers conducted a study examining the dyadic relation between two connected products [22]. They found that, when they converged the eWOM ratings for these products, there was a corresponding increase in the total sales. However, another study defined the "Often bought together" co-product-network-based recommendation as a "complementary" recommender, and found that a link through this type of recommender led to diverging product ratings [20]. In an attempt to uncover the impact of product distance on eWOM, Pan et al. analyzed a book co-purchase network on Amazon. They discovered that a focal product's ratings were influenced by that of its neighboring products, with the number of paths connecting these products being closely related to the similarity of the eWOM [21], whereas there is no empirical evidence that explores the impact of online reviews on the formation of co-purchase links. Simultaneously, managing eWOM in the co-purchase network is a relatively new topic, which has the potential for retailers to improve eWOM and employ it for cross-selling at the dyadic product-to-product level. Moreover, understanding of the directional relationship between the co-purchase network and eWOM valence (such as negative ratings and positive ratings) remains scarce in the literature [20,21]. Hence, this study leverages the complex network analysis to study the association between the consistency of eWOM (via review volume and positive ratings at the dyadic product-to-product level) and the formation of co-purchase links.

2.3. Methodology for Electronic Network Analysis

In the conventional methodology for electronic network analysis, the ultimate objective is different (such as sales or product node sales rank as the response) [7,9,10], or the endogenous network-based effects and the extrinsic factors of the products/nodes are studied independently and separately [12–14]. The prior work cannot effectively uncover the process that shapes the co-purchase network.

The common network modeling methods used are the bipartite network model [29,30] and ERGM [15,31,32]. ERGMs are a family of random graph probabilistic models that draw a statistical inference of the local selection on the processes shaping the formation of the global structure of an observed network [23,24]. These models estimate the interdependencies between network participants and can model complex theoretical networks. On the other hand, bipartite network models are more static and may not accurately represent real-world networks that consider both internal and external factors. The purpose of ERGMs is to parsimoniously describe the local selection forces that shape the global structure of a network [33,34]. Based on the distinction between explanatory models and prediction models, the objective of an explanatory model is to comprehend the influence of independent variables (IVs) on a dependent variable (DV), while that of a prediction model is to predict the DV based on the IVs [35]. ERGMs can be analogous to logistic regression (a type of explanatory model used in empirical research), which predicts the probability that a pair of nodes in a network will have a link between them, taking into account the directionality of the link. Overall, ERGMs are widely used in directed network analysis and the data-driven paradigm to examine the role of IVs on DV, given their interpretability to conduct regression-like analyses. For example, ERGMs have been applied to explain the evolution of trade networks [36,37], the interdependence of stock market returns within a financial network [38], conversational relationships on online social media networks [39], and the formation of disaster management networks [40], as well as the evolution factors of maritime transport networks [41] and the co-interest network of products and brands [15,31,42].

Within the realm of e-commerce, researchers have applied ERGMs to the co-consideration network of products in the Chinese auto market, as inferred from the survey data of buyers. They found that vehicle attributes and similarities in customer demographics are closely linked to the co-consideration between two products [15,42]. Additionally, Xu and Bhattacharyya studied the consumer associative brand network based on Facebook activity data and discovered that consumers tend to be interested in two brands when they belong to the same category, and that the co-engagement relationship is mutual [31]. However, the links in these networks are based on subjective thresholds between products or brands [15,42]. There is a limited understanding of the relationship between the directional formation of co-purchasing and the aggregation of customer preferences and product associations.

This study takes a further step to probe the formation and dissolution of co-purchase links in understanding co-purchase relations by considering eWOM, product categories, and topological structure. We propose an ERGM to address our research questions. This data-driven approach can enhance the results' reliability and uncover new relationships within empirical explanatory model research. There are several advantages of our proposed ERGM framework in contrast to previous studies. First, the ERGM can integrate the internal topological structure effects of network and external factors, which provides a systematic mechanism to explore the factors impacting the formation of the co-purchase network. Second, the proposed approach is a rigorous, longitudinal, network-based, and analytical framework to study the probability of a spillover increase in co-purchase links, given a certain temporal change of each variable among eWOM.

3. Materials and Methods

This section outlines our research methodology framework. Firstly, it describes the structure of clickstream data, including product categories, user-product interactions, time, and online reviews. Secondly, we build a co-purchase network based on the product co-purchase records of the same users in the clickstream data. Then, we derive the indicators of network topology and eWOM. Last, we establish the exponential random graph model (ERGM) framework.

3.1. Data Structure

Our dataset is a panel dataset, also known as clickstream data, containing a large number of user–computer interactions [43,44]. The data structure is shown in Table 1. User 1 first clicked on product A and viewed it for 8 s, then visited product B three times over a period of 5 min and 18 s. Finally, she purchased product C, then product D, and product B. The online review information is mapped by product ID and is included in Table 1. The co-purchase relationships are represented by a link from product C to product D, and another link from product D to product B, as shown in Table 2.

3.2. Co-Purchase Network Construction

The design of directed co-purchase recommendation hyperlinks in e-commerce provides targeted marketing and cross-selling opportunities. These hyperlinks, which connect a focal product to other products, are important for precision marketing and cross-selling tactics [7,8,10]. A co-purchase link not only reflects the demand relationship between two products, but also captures the collective preferences of many users [10,45]. As seen in Tables 1 and 2, this study extracts an adjacency matrix y by connecting products if they are purchased by the same user, and constructs a directed co-purchase network G that reflects the co-purchase relationships based on the collective preferences of users.

Table 1. A clickstream example for a shopping session of User 1.

User ID	Product ID	Positive Ratings [1]	Review Volume [2]	Category	Interaction	Timestamp
1	A	0.974	1	3	Click	2016/3/18 10:00:56
1	B	0.943	0	4	Click	2016/3/18 10:01:04
1	B	0.943	0	4	Click	2016/3/18 10:01:32
1	B	0.943	0	4	Click	2016/3/18 10:02:01
1	C	0.982	1	7	Click	2016/3/18 10:06:22
1	D	0.979	1	6	Click	2016/3/18 10:06:34
1	C	0.982	1	7	Purchase	2016/3/18 10:08:02
1	D	0.979	1	6	Click	2016/3/18 10:08:04
1	B	0.943	0	4	Click	2016/3/18 10:08:43
1	D	0.979	1	6	Purchase	2016/3/18 10:09:56
1	B	0.943	0	4	Purchase	2016/3/18 10:10:02

[1] Positive rating (ratio) of online product review range from 0 to 1. [2] Review volume is a categorical variable, =0 for 0 to 50 reviews, indicating a less popular product; =1 for over 50 reviews, indicating a highly popular product. This is a business setting of this retailer and further details can be found in Section 3.3.2.

Table 2. Illustration of constructing a co-purchase adjacency matrix.

	Adjacency Matrix y			Formation of a Co-Purchase Network G
End j / Start i	Product B	Product C	Product D	
Product B	0	0	0	
Product C	0	0	1	C ⇒ D ⇒ B
Product D	1	0	0	

Formally, the co-purchase network can be defined by the graph $G = \{N, E\}$, consisting of a set of product nodes N and a set of E co-purchase edges that interconnect them. The network $G = \{N, E\}$ is also represented by its adjacency matrix $y = \{y_{ij}\}, i, j = 1, 2, \cdots, n$, which is an $n \times n$ matrix. Wherein,

(1) The (i, j) entry $y_{ij} = 1$ for a link from product i to product j denotes that users have bought product i and then product j; otherwise, $y_{ij} = 0$. We note that y_{ij} is not necessarily equal to y_{ji}, due to the link direction of co-purchase relations.

(2) The diagonal entries are stipulated to be $y_{ij} = 0$, denoting that repurchase behavior towards a product is not the focus of this study.

Simultaneously, for measuring the economic effect of co-purchases, we construct a weighted co-purchase network with a weighted adjacency matrix $W = \{w_{ij}\}$. Analogously,

(1) Each link has a weight w_{ij}, indicating the count of the co-purchases from product i to product j. Each count indicates two purchases in total for product i and product j. We consider the count instead of the sales volume due to the characteristic of co-purchases. w_{ij} is not necessarily equal to w_{ji}.

(2) The diagonal entries $w_{ij} = 0$ denote that repurchases are not considered. $W = \{w_{ij}\}$ measures the directed demand effect in terms of a product's sales, which are lifted by others to a certain extent [8–10].

It should be noted that we assume that the sequence of products purchased, as recorded in the clickstream data by the e-retailer, is a reasonable approximation of the interdependency of one product to another. Based on this assumption, we can establish connections in an adjacency matrix that represents the co-purchasing relationship of products within the same online basket. To construct the co-purchase network, our clickstream data are processed as follows:

Step 1: Obtaining individual user-level purchase records. We obtained these purchase records grouped by user ID and arranged them chronologically from the clickstream data (e.g., a user's purchase records as shown in Table 1). To focus on the co-purchase relationship, we removed the records with only one product.

Step 2.: Identification of collective co-purchase relations. Using an enormous number of individual users' purchase records, which were obtained by step 1, we followed the work of ref. [45] and exploited a co-occurrence algorithm [46] to identify the collective co-purchase relations. The matrix $y = \{y_{ij}\}$ (e.g., Table 2 displays its formation on the right side) and the weighted one $W = \{w_{ij}\}$ can be constructed.

Step 3: Visualizing the co-purchase network. The co-purchase network, including its weighted version, can be represented through network visualization tools (such as Ucinet [47]).

3.3. Statistics Measures

Various statistical attributes can be extracted from the clickstream data, including both internal network characteristics (such as degree distribution) and eWOM factors (such as online product reviews). These metrics are used to provide a comprehensive understanding of the formation of co-purchasing for the products within the observed network.

3.3.1. Network Topological Attributes

Out-degree centrality: The out-degree centrality of product i refers to the number of outgoing co-purchase links originating from product i within the adjacency matrix $y = \{y_{ij}\}$:

$$d_i^{out} = \sum_j y_{ij} \tag{1}$$

In-degree centrality: The in-degree centrality of product i refers to the number of incoming co-purchase links to product i within the adjacency matrix $y = \{y_{ij}\}$:

$$d_i^{in} = \sum_j y_{ji} \tag{2}$$

It is noted that each link has a weight (w_{ij}), which represents the count of the co-purchases between product i and product j. Each count denotes two purchases in total for both products. As a result, each count is equivalent to half the sales volume. This study uses the count of the co-purchases as a proxy for sales volume due to the nature of co-purchases. To measure the strength of the sales volume from product i's outgoing co-purchase links and incoming co-purchase links, we define the out-strength and in-strength of product i, respectively.

Out-strength: The out-strength of product i refers to the total count of outgoing co-purchases originating from product i within the weighted adjacency matrix $W = \{w_{ij}\}$:

$$s_i^{out} = \sum_j w_{ij} \tag{3}$$

In-strength: The in-strength of product i refers to the total count of incoming co-purchases targeting product i within the weighted adjacency matrix $W = \{w_{ij}\}$:

$$s_i^{in} = \sum_j w_{ji} \tag{4}$$

3.3.2. eWOM Factors

The volume and valence of eWOM provide potential consumers with valuable information for their decision making process [18,19,28]. The focal e-retailer rigorously provides types of online product review volumes and ratings. The eWOM management mechanism is implemented in e-commerce retail information systems and websites. Specifically, a five-star rating system that is used by the focal e-retailer allows for customers to rate their feedback on a scale from 1 to 5. A negative review is identified as a product review where the customer selects 1 star when evaluating a product they purchased. The more stars selected, the more positively the customer is responding to the product in their review.

This study utilizes the volume of online reviews to assess the impact of eWOM volume (categorical variable, rv_i). Positive ratings (percentage variable, pr_i) are employed to measure the valence of eWOM.

eWOM Volume: The categorical variable (rv_i) represents the review volume of product i, which is accessible at time t:

$$rv_i = \begin{cases} 0, \text{ 1–50 pieces of reviews} \\ 1, \text{ Over 50 pieces of reviews} \end{cases} \tag{5}$$

A product with over 50 reviews simply means that it is much more popular compared to products with fewer reviews, which is a business setting of the e-retailer. However, this setup is reasonable for two reasons. First, in this paper, a product refers to a stock-keeping unit (SKU). SKUs are unique codes assigned to each distinct product and its variations within a retailer's inventory system. Each SKU represents the smallest unit of a product that can be sold, tracked, and managed in an inventory. For example, a piece of clothing may have three SKUs: white, red, and black. Therefore, it is reasonable to regard an SKU product as relatively popular if it has more than 50 reviews. Second, the e-retailer has a product delisting policy. For example, if a product, such as a book, mobile carrier service, or baby product, has less than 50 reviews, and if its proportion of negative reviews (i.e., 1-star reviews) is greater than 9%, then there is a high likelihood that the e-retailer will delist the product. This delisting policy is in place to ensure the quality of the retailer's product offerings and to protect the interests of its customers.

For feasible management practices, we employ these mechanisms without altering the volume and valence of eWOM. This enables the e-retailer to efficiently implement management strategies based on our research findings, without undergoing extensive modifications to the current eWOM management system.

eWOM Valence: The positive rating (pr_i) of online product reviews is expressed as 1 minus the ratio of the number of negative reviews to the total review volume for product i, which is accessible at time t:

$$pr_i = 1 - \frac{\text{Number of negative reviews } (i)}{\text{Total number of reviews } (i)} \tag{6}$$

3.4. Modeling the Formation of Co-Purchase Network

3.4.1. Exponential Random Graph Modeling

The purpose of using the ERGM framework is to estimate the ERGMs to decide the most parsimonious one, and then to examine the impacts of the change of local network configurations on the formation of the global structure of the co-purchase network. This work aims to study the impacts of online WOM on the formation of co-purchase links between products. Our observed co-purchase network $G = \{N, E\}$ could be considered to be like the outcome of a traditional regression model, where the likelihood of consumers consecutively purchasing two products can be formulated as the probability of a co-purchase link. However, apart from the regression approach and its independence assumption, the ERGM models the network by assuming a complex conditional dependence (refer [23,24]). Specifically, the ERGM framework has a Markov dependency, which indicates that, when the value of a link changes (e.g., within the adjacency matrix $y = \{y_{ij}\}$, another expression of the observed co-purchase network $G = \{N, E\}$, e.g., a link from product i to product j toggle states, such as $y_{ij} = 1 \Rightarrow y_{ij} = 0$), the probability of the other links will change (even if the rest of the network keeps constant).

We define the set of all random networks and the observed network as Y and y in form of an adjacency matrix, correspondingly. In this study, Y is a subset of all the $n \times n$ asymmetric matrices, since there are n products in the observed directed co-purchase network. The ERGM specifies the probability of the observed network y, following the distribution of Y, which can be expressed as:

$$P_{\theta,Y}(Y = y) = \frac{\exp\{\theta^T g(y, X)\}}{K(\theta, Y)}, y \in Y \tag{7}$$

Wherein, (i) θ represents the vector of the model coefficients that estimates the magnitude of the effect of each variable; (ii) $g(y, X)$ is a q-vector of the statistics containing the topological structure, product category, and online WOM factors for individual product nodes, conditional on a matrix y; and (iii) its denominator $K(\theta, Y)$ is the normalizing term, ensuring that Equation (7) is a legitimate probability function over all the possible networks y.

Equivalently, owing to the Markov dependency in the ERGM, Equation (8) states that the log-odd that any given co-purchase link will exist, given the current state of the reminder of all other links in this network, is:

$$\text{logit}(Y_{ij} = 1) = \theta^T \delta[g(y, X)]_{ij} \tag{8}$$

In Equation (8), Y_{ij} is a directed link in a random Y and $\delta[g(y, X)]_{ij}$ is the change in $g(y, X)$ if the value of y_{ij} is switched from 0 to 1. The Markov chain Monte Carlo maximum likelihood estimation (MCMCMLE) method is employed to estimate the parameters [33,34]. The log-likelihood and Akaike information criterion (AIC) are two standard measures of model fitness to select the most parsimonious ERGM.

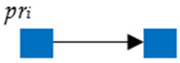
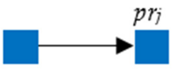
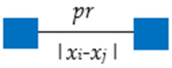
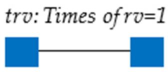
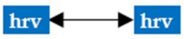
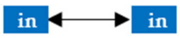
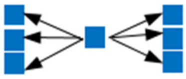
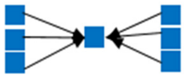
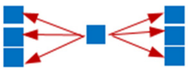
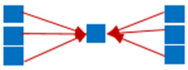
3.4.2. Local Configuration of ERGM Variables

We adopt an ERGM framework to identify whether online reviews, product categories, and topological structures are significant factors in impacting the co-purchase formation. Equation (8) in the framework shows that the response variable is the log-odds of forming a co-purchase link. Table 3 shows the configuration and definition of the explanatory variables used in the ERGM model. It is worth noting that the observed co-purchase network is directed. Thus, the local configurations at the product-to-product level in Table 3 enable us to study the product demand interdependence beyond just understanding collective co-purchase preferences.

Table 3. Construction and definition of EGRM explanatory variables.

Variable	Configuration	Statistic	Definition
Edge term	$\blacksquare \rightarrow \blacksquare$	$\sum_{i \neq j} y_{ij}$	The number of co-purchase links in the network.
Category: #1	$1 \leftrightarrow 1$	$\sum_{i \neq j} y_{ij} \delta_{ij}^{\text{cate1}}$	The number of links (i,j) for both product i and product j belong to product category #1. If both belong to product category #1, $\delta_{ij}^{\text{cate1}} = 1$; $\delta_{ij}^{\text{cate1}} = 0$, otherwise.
Category: #2	$2 \leftrightarrow 2$	$\sum_{i \neq j} y_{ij} \delta_{ij}^{\text{cate2}}$	The number of links (i,j) for both product i and product j belong to product category #2. If both belong to product category #2, $\delta_{ij}^{\text{cate2}} = 1$; $\delta_{ij}^{\text{cate2}} = 0$, otherwise.
Category: #3	$3 \leftrightarrow 3$	$\sum_{i \neq j} y_{ij} \delta_{ij}^{\text{cate3}}$	The number of links (i,j) for both product i and product j belong to product category #3. If both belong to product category #3, $\delta_{ij}^{\text{cate3}} = 1$; $\delta_{ij}^{\text{cate3}} = 0$, otherwise.
Category: #4	$4 \leftrightarrow 4$	$\sum_{i \neq j} y_{ij} \delta_{ij}^{\text{cate4}}$	The number of links (i,j) for both product i and product j belong to product category #4. If both belong to product category #4, $\delta_{ij}^{\text{cate4}} = 1$; $\delta_{ij}^{\text{cate4}} = 0$, otherwise.
Category: #5	$5 \leftrightarrow 5$	$\sum_{i \neq j} y_{ij} \delta_{ij}^{\text{cate5}}$	The number of links (i,j) for both product i and product j belong to product category #5. If both belong to product category #5, $\delta_{ij}^{\text{cate5}} = 1$; $\delta_{ij}^{\text{cate5}} = 0$, otherwise.

Table 3. *Cont.*

Variable	Configuration	Statistic	Definition
Category: #6		$\sum_{i \neq j} y_{ij}\delta_{ij}^{\text{cate6}}$	The number of links (*i*,*j*) for both product *i* and product *j* belong to product category #6. If both belong to product category #6, $\delta_{ij}^{\text{cate6}} = 1$; $\delta_{ij}^{\text{cate6}} = 0$, otherwise.
Category: #7		$\sum_{i \neq j} y_{ij}\delta_{ij}^{\text{cate7}}$	The number of links (*i*,*j*) for both product *i* and product *j* belong to product category #7. If both belong to product category #7, $\delta_{ij}^{\text{cate7}} = 1$; $\delta_{ij}^{\text{cate7}} = 0$, otherwise.
(a) Positive ratings for outgoing products		$\sum_{i} y_{ij} pr_i$	The total value of the positive ratings of product *i* for all outgoing links (*i*,*j*).
(b) Positive ratings for incoming products		$\sum_{j} y_{ij} pr_j$	The total value of the positive ratings of product *j* for all incoming links (*i*,*j*).
(c) Dyadic difference of positive ratings		$\sum_{i \neq j} y_{ij}\lvert x_i = x_j \rvert,$ $x = pr$	The sum of absolute difference of the positive ratings of product *i* and product *j* for all links (*i*,*j*).
(d) Factor attribute effect of review volume with over 50 reviews (reference: 0–50 reviews).		$\sum_{i \neq j} y_{ij} trvi_{i,j}$	The number of times products with over 50 reviews appear in a link $trvi_{i,j}$ for all links (*i*,*j*).
(e) Homophily of review volume with over 50 reviews (reference: 0–50 reviews).		$\sum_{i \neq j} y_{ij} hrvi_{i,j}$	The number of links (*i*,*j*) for both product *i* and product *j* with over 50 reviews. If both have over 50 reviews, $hrvi_{i,j} = 1$; $hrvi_{i,j} = 0$ otherwise.
(f) Interaction effect of positive ratings and review volume.		$\sum_{i \neq j} y_{ij}(x_i + x_j),$ $x = in$	The sum of product *i*'s interaction $in_i = pr_i \times rv_i$ and product *j*'s $in_j = pr_j \times rv_j$ for all links (*i*,*j*).
(g) Out-degree centrality		$\sum_{i} y_{ij} d_i^{out}$	Out-degree centrality of product *i* in all outgoing links (*i*,*j*).
(h) In-degree centrality		$\sum_{j} y_{ij} d_j^{in}$	In-degree centrality of product *j* in all incoming links (*i*,*j*).
(i) Out-strength		$\sum_{i} y_{ij} s_i^{out}$	Out-strength of product *i* in all outgoing links (*i*,*j*).
(j) In-strength		$\sum_{j} y_{ij} s_{in}^{j}$	In-strength of product *j* in all incoming links (*i*,*j*).

Note: $\lvert x_i - x_j \rvert$ denotes the absolute-value norm on the 1-dimension space.

In Table 3, the Edge term represents the general propensity of the network to form co-purchase links. It is worth noting that the Edge term is not one of the three key factors that we examined in this study. In ERGM analyses, the Edge term is a basic term similar to an intercept in a regression model. This statistic counts the number of co-purchase links in the network, in order to interpret the baseline likelihood for a product node to form a co-purchase link with another one. In co-purchase recommendation systems, the value of the links should be generalized across the different product categories to balance the sales increase and diversity [4,8,25]. Product category is included in the ERGM model to gain insight into consumers' cross-purchasing behavior across categories. There are seven main product categories in e-retailers. The differential homophily effect of the multi-category variables in the ERGM describes each group within the variables as having a unique propensity for within-group ties [34,48]. To examine the differential homophily effect of product category, this study allows each product category to have a propensity for within-category co-purchase links, as measured, respectively, in Table 3.

The external eWOM factors (in the form of online product reviews) are also constructed at a product-to-product level, as shown in Table 3. For the eWOM valence, variable (a) is used to test the effect of the positive ratings for the products in outgoing co-purchase links, while variable (b) assesses the impact of the positive ratings on the incoming co-purchase link formation. We also examine the homophily effect of the positive ratings by constructing variable (c), which measures the dyadic difference of the positive ratings between two connected products. When examining the eWOM volume, variable (d) is utilized to evaluate the factor-attribute impact of the review volumes with more than 50 reviews. The homophily effect of the review volume is also measured by variable (e). The reference level is set as a review volume with 0–50 reviews. This is because the inclusion of all levels would lead to a linear dependency, which should be avoided here, akin to all statistical models [33,34]. Furthermore, the study regulates the interaction effect of positive ratings and review volume with variable (f).

The network topological attributes, (g) to (j), are referred to as covariates and are used to examine the scale-free structure in the co-purchase network. These covariates assess the asymmetric bilateral relationship and examine the tendency for outgoing products to diversify and incoming products to concentrate. In particular, variables (i) and (j) measure the strength of the sales volume.

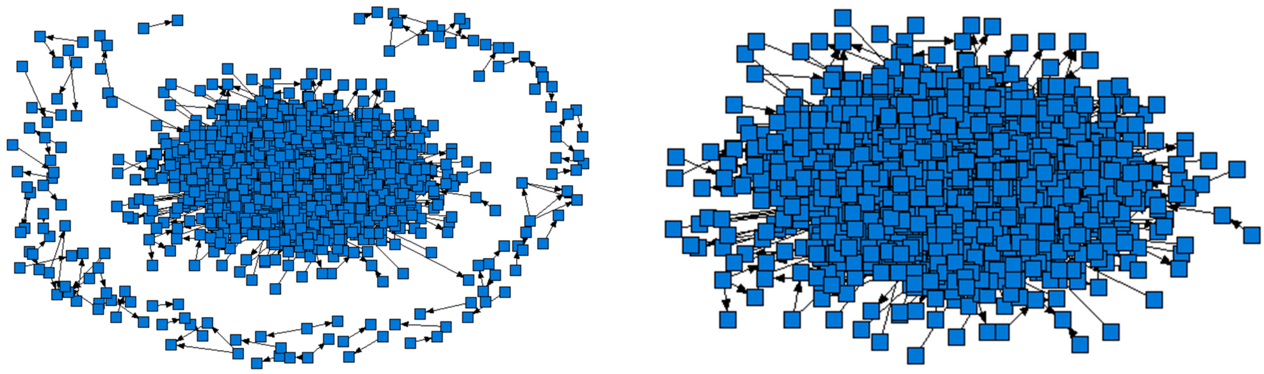
4. Experiments and Empirical Results

In this section, we first provide a description of the dataset and variables. Then, we analyze the topology properties of the co-purchase network from both global and local perspectives. Lastly, we provide ERGM diagnostics and estimation for discussion about the factors that affect the formation of the observed co-purchase network.

4.1. Description of Dataset and Basic Variables

Our dataset comes from the web logs of a leading Chinese e-retailer that offers a vast array of product categories, such as electronics, computers, home appliances, and books, from March 1 to March 31 in 2016. A total of 83,884 users viewed the pages of 22,364 unique products, resulting in 6,759,711 user–product interactions and 19,233 transactions. The time window is not a shopping holiday like Black Friday, Singles Day, Double Twelve, or the period of Chinese New Year in February, when express delivery services are suspended. The data were provided for collaborative purposes without a potential impact from the recommendation system, to seek better personalized recommendation solutions, and to achieve a higher recommendation efficiency. Festival marketing is not the focus of this study and will be the subject of our next research. To analyze the formation of co-purchase networks, we used a co-occurrence algorithm to convert the clickstream data into an adjacency matrix [46]. This resulted in 1825 co-purchased products (8.16% of total), making up 73.7% of the total transactions, as shown in Figure 3, demonstrating a long tail phenomenon with a few hit products that dominated the sales [5,49].

Table 4 provides the statistical description of the basic variables. We found that the range of the out-degree centrality was larger than that of the in-degree centrality, while the range of the out-strength was smaller than that of the in-strength. The distributions of the degree centrality and strengths had similar heavy tails, with 95% of the out-degree (in-degree) centrality ranging from 0 to 14, and 95% of the out-strength (in-strength) ranging from 0 to 17 (0 to 18). However, their standard deviation was much bigger than their average. This indicates that there is heterogeneity in consumers' co-purchase behaviors. When it came to eWOM, we observed that the average positive rating of the online reviews was 0.96 and 86.91% out of the 1825 products had over 50 reviews. This indicates that the majority of consumers were pleased with the products in the co-purchase network and that a large proportion of these products were highly sought after in the e-retailer. Additionally, the co-purchase network consisted of 7 product categories, 1825 product nodes, and 5731 co-purchase links. The distribution of product categories is nearly uniform, with the exception of category #7 having a lower likelihood.



(**a**) (**b**)

Figure 3. The observed co-purchase network: (**a**) the global topology using the adjacent matrix $y = \{y_{ij}\}$; and (**b**) the main component using the weighted adjacent matrix $W = \{w_{ij}\}$ after removing the links whose link weights $w_{ij} < 4$.

Table 4. Descriptive summary for basic variables.

Construct		Basic Variables	Mean	SD	Min	Max
Topological structure		Out-degree centrality d_i^{out}	3.14	5.34	0	66
		In-degree centrality d_i^{in}	3.14	5.55	0	48
		Out-strength s_i^{out}	3.88	8.69	0	145
		In-strength s_i^{in}	3.88	9.04	0	185
eWOM	eWOM valence	Positive ratings pr_i	0.96	0.03	0	1
	eWOM volume	Review volume rv_i	=0 for 0–50 reviews: 239 (0.13); =1 for over 50 reviews: 1586 (0.87).			
Product category			Category #1: 335 (0.18); Category #2: 287 (0.16); Category #3: 292 (0.16); Category #4: 339 (0.19); Category #5: 281 (0.15); Category #6: 260 (0.14); and Category #7: 31 (0.02).			

4.2. Global Structure of the Co-Purchase Network

To statistically compare the out-degree centrality and in-degree centrality, as well as the out-strength and in-strength, from the perspective of a global network structure, we use the power-law model $P(x) = \int_x^\infty p(x')\, dx' = \left(\frac{x}{x_{\min}}\right)^{-\alpha+1}$ to fit the distributions of d_i^{out}, d_i^{in}, s_i^{out}, and s_i^{in}. Specifically, we propose a hypothesis to test the goodness of the power-law model fit by using the Kolmogorov–Smirnov (KS) statistics via a bootstrapping procedure (refer [50]):

H_0: *The data of a network attribute follows a power-law distribution.*

H_1: *The data of a network attribute does not follow a power-law distribution.*

Table 5 presents the results of the power-law fits and the tests for the data of the network topological attributes. The estimated exponents α and cut-offs $x_{\min}$, which are shown on the left of Table 5, suggest that the degree centrality and strengths are right-skewed distributions. The power-law test indicates that there is no reason to reject the null hypothesis H_0, except for the in-degree centrality (its p-value is less than 0.05). These tests confirm that the observed co-purchase network is a typical small-world network, where most products have fewer co-purchase links and less sales volume, and a small number of products have a majority of co-purchase links and more sales volume. These results validate the Pareto principle in product sales from a perspective of network science [4,49].

Table 5. Summary of power-law fits and tests for network topological attributes.

Variable	Power-Law Fit			Power-Law Test *			
	Exponent α	Cut-off $x_{\min}$	Std KS	Std $x_{\min}$	Std α	Std Tail	p-Value
Out-degree centrality d_i^{out}	4.999	26	0.016	6.233	1.257	32.481	0.73
In-degree centrality d_i^{in}	3.242	13	0.010	1.961	0.353	21.042	0.00
Out-strength s_i^{out}	3.107	20	0.012	5.195	0.390	29.956	0.21
In-strength s_i^{in}	3.654	28	0.014	7.984	0.629	42.252	0.34

* Notes that the power-law test uses bootstrapping procedure depending on 2000 simulations.

4.3. Concentration Effect of the Co-Purchase Network

A product with a high in-degree centrality indicates its popularity and can drive more traffic to its webpage by being sold with many other products. On the other hand, a product with a high out-degree centrality means that it has a larger number of other products that are directly accessible from its webpage, which may indicate a relationship with original products and their accessories. Figure 4 shows the correlation between these degree centralities. We observe a smaller variation of the in-degree centrality across the products compared with the out-degree centrality, but both distributions are similar, as seen in Table 4 and Figure 4. However, Table 5 shows that the data of the in-degree centrality do not follow a power-law distribution from a statistical inference perspective. To examine the intrinsic concentration effect of product node connectivity, or the mutuality effect, we conducted a Pearson correlation test and found that the in-degree centrality and out-degree centrality were positively correlated (correlation value = 0.28, p-value = $2.2 \times 10^{-16} < 0.001$). These results imply the existence of an intrinsic concentration effect in the network, where products have mutual co-purchase relations.

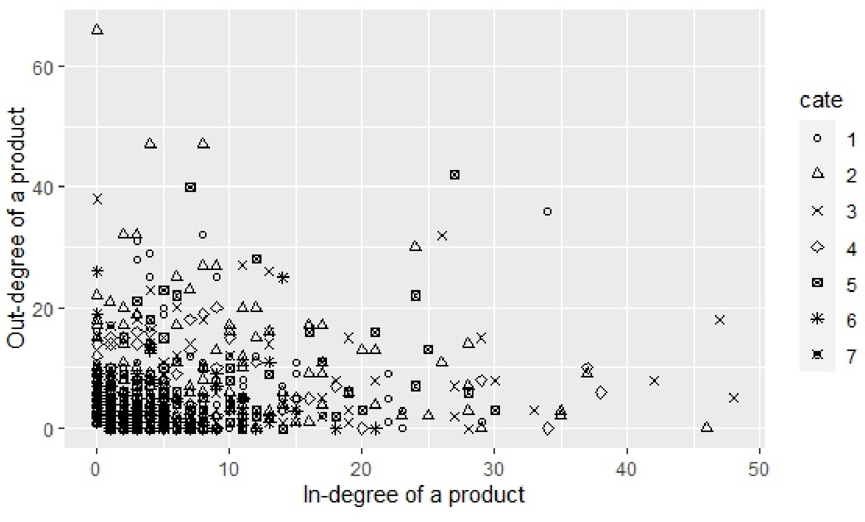


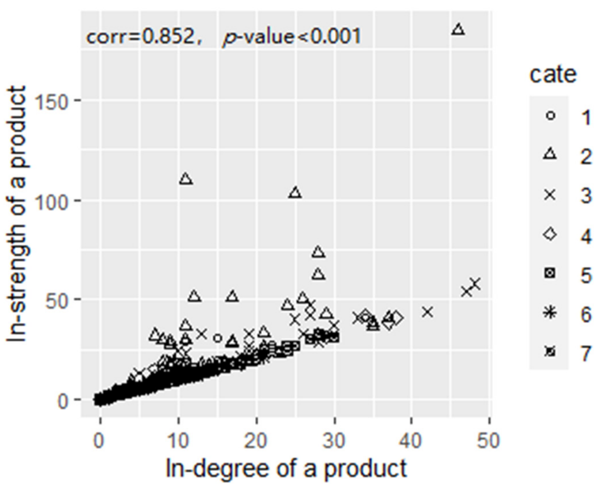
Figure 4. Plot of the correlation between in-degree centrality and out-degree centrality.

Besides the intrinsic concentration effect, we aim to study the relationship between a product's aggregate sales volume and its co-purchasability. The concept of homophily suggests that consumers tend to purchase products that are similar to those that they have previously purchased [1,2]. This phenomenon is known as co-purchasing and has been extensively studied in network research [13]. To test the potential directionality in a co-purchase network, the degree centralities represent the extent to which consumers purchased a product i with other products (i.e., product i's co-purchasability), and the strengths indicate the aggregate sales volume of product i within its shared co-purchase relations (i.e., the strength of product i's connected co-purchase links). Prior studies have also shown that the degree centrality of a node (in this case, a product) is a good indicator of its importance within the network [15,31,42]. The degree centrality represents the number of connections a node has with other nodes. Previous research has found that the sales volume of a product is correlated with degree centrality [8–10]. In other words, products that have a higher degree centrality are likely to be more important within the co-purchase network. Therefore, it is reasonable to hypothesize that products with a higher degree centrality (i.e., more co-purchases) will also have a higher total sales volume. Specifically, we posit that a product's co-purchasability is positively correlated with the aggregate sales volume between itself and its connected products. To examine this hypothesis, we performed a Pearson correlation test:

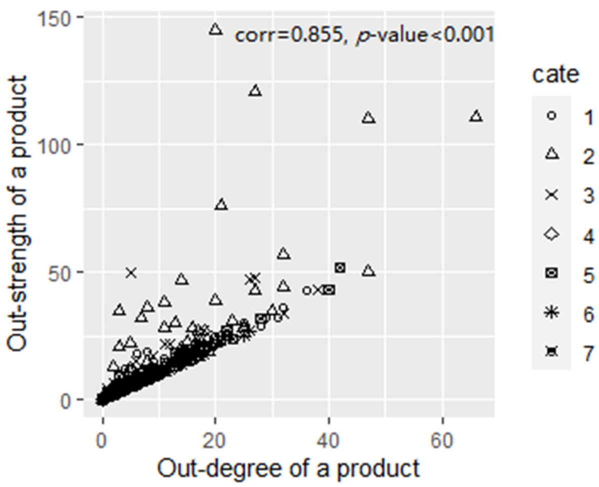
H_0: *The in-degree centrality (out-degree centrality) of a product is independent of its in-strength (out-strength).*

H_1: *The in-degree centrality (out-degree centrality) of a product is correlated with its in-strength (out-strength).*

As shown in Figure 5, we can see a close linear relationship between the degree centralities and strengths. Through a Pearson correlation analysis, we found that our hypotheses H_1 are supported in terms of the incoming links (correlation value = 0.852, p-value < 0.001 in Figure 5a) and outgoing links (correlation value = 0.855, p-value < 0.001 in Figure 5b), respectively. This means that products with higher degree centralities have larger aggregate sales volumes, which demonstrates the directed rich-get-richer concentration effect in co-purchase behavior. The network effects in online markets have shown that products with a higher centrality tend to have a higher sales volume [8–10]. These studies suggest that products with a higher centrality benefit from a rich-get-richer effect, where they attract more customers due to their prominent position in the network. Furthermore, we analyzed the directed relationship between the degree centrality and sales volume in the context of a co-purchase network, which focuses on co-consideration and co-purchasing instead of co-engagement [15,31,42]. The findings enhance our understanding of the directed rich-get-richer effect by examining the dyadic relationships between the products in terms of their co-purchase behavior.



(**a**)



(**b**)

Figure 5. Rich-get-richer concentration effect test: (**a**) incoming perspective; and (**b**) outgoing perspective.

We also note that the product category affects both the intrinsic and rich-get-richer concentration effects. For example, the products belonging to the second category, shown as triangle-shaped scatters in Figures 4 and 5, have more intense concentration effects, as these scatters are distributed above the averages (as indicated by the slope of the approximate linear regressions) in Figure 5a,b. In the study, we control for the effect of the product category in the ERGM diagnostics and estimation.

4.4. ERGM Diagnostics and Estimation

In total, three ERGM models are proposed with different specifications under a paradigm of hierarchical regression-like analysis [34,42,51]. Model A is a basic model that only includes two variables, edges (explaining density) and product category (explaining cross-category purchase patterns), Model B is an extended model that adds the network topological attributes to investigate the bilateral relationship, and Model C is a refined model that includes all the variables and emphasizes the impact of eWOM. The fitting process is based on the MCMCMLE method, and several model selection criteria, including log-likelihood and AIC (Akaike information criterion), are used to select the most parsimonious model [33,34].

The fitting processes of Model A to Model C are convergent. According to the model selection criteria [33,34] and the similar research process [37,42], the reduced AIC value and bigger log-likelihood value from Model A to Model C in Table 6, indicate that the improvement of the model fitness is due to the addition of the network topological structure and eWOM. Model C is selected as the best-fitted model, with the maximum log-likelihood (−41,538.06) and the minimum AIC (83,112). Furthermore, empirical research considers the coefficient estimation to be robust if the signs of the parameters are consistent in the hierarchical regression analyses [42,51]. The proposed ERGM is a data-augmentation explanatory model, due to its interpretability to conduct regression-like analyses [34,42]. In Table 6, we observe that both the signs and the significance level of the estimates are consistent from Models A to C, suggesting that the impacts of the topological structure, product category, and online reviews on co-purchase are robust. Overall, Model C is considered to be the best fitted model to study the factors that affect the formation of the co-purchase network.

Table 6. Results of the ERGMs' fitting for the formation of the co-purchase network.

Model	DV: Log-Odds of a Co-Purchase Link Formation.		
	A	B	C
Edge term	−6.371 *** (0.015)	−6.213 *** (0.020)	−3.671 *** (1.241)
Product category			
Category: #1	0.332 *** (0.063)	0.330 *** (0.063)	0.319 *** (0.063)
Category: #2	−0.008 (0.086)	−0.063 (0.087)	−0.061 (0.087)
Category: #3	−0.036 (0.086)	0.059 (0.086)	0.099 (0.086)
Category: #4	−0.159 ** (0.079)	−0.191 ** (0.079)	−0.253 *** (0.080)
Category: #5	0.248 *** (0.078)	0.272 *** (0.078)	0.280 *** (0.078)
Category: #6	−0.326 *** (0.111)	−0.340 *** (0.111)	−0.335 *** (0.111)
Category: #7	0.638 (0.578)	0.570 (0.579)	0.602(0.579)
eWOM			
(a) Positive ratings for outgoing products	-	-	−2.449 *** (0.678)
(b) Positive ratings for incoming products			−0.076 (0.736)
(c) Dyadic difference of positive ratings	-	-	−3.098 *** (0.683)
(d) Factor attribute effect of review volume with over 50 reviews (reference: 0–50 reviews).	-	-	2.793 *** (0.663)
(e) Homophily of review volume with over 50 reviews (reference: 0–50 reviews).	-	-	−0.095 (0.059)
(f) Interaction effect of positive ratings and review volume	-	-	−2.862 *** (0.679)
Network topological structure			
(g) Out-degree centrality	-	0.037 *** (0.003)	0.038 *** (0.003)
(h) In-degree centrality	-	−0.095 *** (0.005)	−0.093 *** (0.005)
(i) Out-strength	-	−0.004 * (0.002)	−0.004 * (0.002)
(j) In-strength	-	0.013 *** (0.003)	0.012 *** (0.003)
Model diagnostics			
Log-likelihood	−42,175.37	−41,573.3	−41,538.06
AIC	84,367	83,171	83,112

Note: Model A only considers Edge term and product category; Model B considers Edge term, product category, and network topological attributes; and Model C considers all variables. Estimate (standard error, S.E.); * p-value < 0.1, ** p-value < 0.05, and *** p-value < 0.01. AIC, smaller is better; and log-likelihood, bigger is better.

Edge term. The Edge term in our directed co-purchase network refers to the connections between the nodes. The network has 1825 nodes, meaning that it has the potential to reach a total of $1825 \times (1825 - 1) \times 2$ = 6,657,600 directed links. In reality, it only has 5731 links, resulting in a density of 0.086% (5731/6,657,600). This suggests that the network is sparse. As demonstrated in Table 6, Model C's coefficient estimate reveals a negative and significant value of −3.671 (p-value < 0.01) for the Edge term. This significance suggests that co-purchase relationships are not likely to occur randomly between the products. Instead, it reflects the genuine preferences and associations among the consumers and products.

Cross-category co-purchase. Additionally, we observe differential homophily effects in consumer cross-category purchasing behaviors. Our results indicate that, when consumers purchase products in categories #1 (0.319, p-value < 0.01) or #5 (0.280, p-value < 0.01), they tend to subsequently purchase products within the same category. On the other hand, if a product belongs to categories #4 (−0.253, p-value < 0.01) or #6 (−0.335, p-value < 0.01), then consumers are more likely to co-purchase it with a product from a different category.

To further investigate this co-purchasing behavior across the categories, we developed a new model based on Model C. This new model replaced the 7 product category variables with 49 cross-category variables (7×7), while keeping all the other settings the same. The significant results are presented in Table 7. While category #1 exhibits a positive homophily effect, Table 7a shows that the consumers who purchase products in category #1 are also likely to purchase products in categories #3 and #6. However, it is not advisable to recommend category #2. Looking at the negative homophily effect, it appears that product category #6 is unlikely to be co-purchased within the same category (as seen in Table 6). However, it is more likely to be purchased together with categories #1, #2, #3, and #5 (as seen in Table 7a). In addition, cross-category purchase behavior is observed from categories that do not exhibit homophily effects, as presented in Table 7b. Thus, it can be concluded that categories #2, #3, and #7 can also significantly impact the co-purchasing behavior across the categories.

Table 7. Impacts of product cross-category purchase on the formation of the co-purchase network.

(a)		**(b)**	
From (with Homophily Effect) -> To	**Estimate (S.E.)**	**From (without Homophily Effect) -> To**	**Estimate (S.E.)**
Category: #1 -> Category: #2	−2.054 *** (0.782)	Category: #2 -> Category: #1	2.281 *** (0.782)
Category: #1 -> Category: #3	1.170 * (0.703)	Category: #2 -> Category: #3	1.855 ** (0.792)
Category: #1 -> Category: #6	1.529 * (0.745)	Category: #2 -> Category: #4	1.715 ** (0.771)
Category: #4 -> Category: #2	−1.679 ** (0.770)	Category: #2 -> Category: #5	1.652 ** (0.784)
Category: #4 -> Category: #6	1.238 * (0.732)	Category: #2 -> Category: #6	2.216 *** (0.829)
Category: #5 -> Category: #1	1.236 * (0.694)	Category: #3 -> Category: #1	1.351 * (0.703)
Category: #5 -> Category: #2	1.953 ** (0.782)	Category: #3 -> Category: #2	1.956 ** (0.791)
Category: #5 -> Category: #6	1.361 * (0.746)	Category: #3 -> Category: #6	1.417 * (0.755)
Category: #6 -> Category: #1	1.724 ** (0.744)	Category: #7 -> Category: #2	1.299 ** (0.505)
Category: #6 -> Category: #2	2.272 *** (0.828)	Category: #7 -> Category: #6	0.908 ** (0.449)
Category: #6 -> Category: #3	1.542 ** (0.690)		
Category: #6 -> Category: #5	1.230 * (0.747)		

Note: * p-value < 0.1, ** p-value < 0.05, and *** p-value < 0.01.

eWOM. As Table 6 demonstrates, positive ratings and review volume were found to be key factors that impact co-purchase formation, owing to their high value of estimates. We found that the positive ratings for outgoing products had a negative impact on the formation of co-purchase links (variable (a), its estimate = −2.449, p-value < 0.01), while the impact of the positive ratings for incoming products was insignificant (variable (b)). The results indicate that a product with a higher positive rating is less likely to subsequently have co-purchased links with other products. eWOM valences, such as positive ratings or negative ratings, reflect consumers' opinions from a global audience [18,52]. An increment of negative publicity about a product often reduces the satisfaction or product attitude of potential consumers [53,54]. Conversely, a highly positive rating of a product increases its sales, which helps consumers to build a perception of a product's quality [55]. Therefore, consumers' acceptable perception of high product quality can be deduced. When a user processes online review information and purchases a product with a highly positive rating, their demand for high-quality products might be met. This implies a lower probability of follow-up purchases for a relatively low-quality product. However, the estimate of variable (b) shows that purchasing a product does not significantly affect the subsequent product (with a highly positive rating) purchases.

Furthermore, if all the other variables are constant, a product with twice the positive rating of another product will have 3.098 lower log-odds of co-purchase, compared with that of two products with the nearly same positive ratings (variable (c), its estimate = -3.098, p-value < 0.01). In other words, a higher inconsistency of positive ratings among products decreases the likelihood of co-purchase. This is comparable to the self-reference effect in information decisions [56,57], where individuals have internalized information about the acceptable positive ratings related to themselves, and they react quicker to products with similar ratings. The inconsistency effect of positive ratings also has the strongest explanatory power, as reflected by its maximum estimated value. The estimation result highlights that the inconsistency effect of positive eWOM ratings can be leveraged to design directed recommendation hyperlinks or sponsored product hyperlinks in e-commerce.

When considering the impact of the eWOM volume, an increased number of online reviews corresponds with a greater product popularity and bolsters consumer buying inclination [19,20,52]. Compared with products with a low popularity (i.e., with 0–50 reviews), the analysis of the variables (d) to (e) reveals that products with a high popularity (i.e., with more than 50 reviews) are more likely to form a co-purchase relationship with another product (2.793, p-value < 0.01). However, there is no clear indication that positive ratings lead to a homophily effect where products with over 50 reviews are more likely to be purchased together. For the interaction effect (variable (g)), we observe that in highly popular products (with over 50 reviews), the higher the positive rating, the lower the likelihood of forming a co-purchase. One possible explanation for this is that consumers find it more difficult to perceive the review's usefulness when the product has many positive reviews [58,59]. Thus, when a product has a higher positive rating and more reviews, consumers may be less likely to seek out additional products to co-purchase with it.

Network topological attribute. The refined Model C incorporates network topological attributes (g) to (j), which show that the number of outgoing connections (out-degree centrality) of a product has a positive effect, whereas the number of incoming connections (in-degree centrality) has a negative effect on the formation of co-purchase links. Despite the negative impact of in-degree centrality, the volume of co-purchases (in-strength) still has a positive, but less significant effect, on the formation of the co-purchase network, whereas the volume of outgoing connections (out-strength) has a trivial effect. These results again suggest that the observed co-purchase network is shaped by both consumer behavior preferences and product association effects.

5. Discussion and Implications

5.1. Discussion

E-commerce platforms benefit from the co-purchase recommendation hyperlinks that are under the label "customers who bought this also bought . . . ". These co-purchase relationships among products reveal consumers' demand dependency on items, indicating that similar consumer groups are likely to respond comparably [1,2]. However, the size of product sets that can be revealed solely by post hoc co-purchases (i.e., historical product purchase records) is limited. This restricts users to a "filter bubble" and has a negative impact on sales diversity [4,6]. In other words, traditional co-purchase recommendation hyperlinks cannot keep up with the evolving demands of consumers. Understanding the formation of the hidden co-purchase relationships between products is an effective approach to optimizing the product allocation and bundle arrangement within recommendation hyperlinks for sales revenue and cross-selling. This study aims to identify the factors that affect the formation of co-purchases, including both internal and external factors. Specifically, we examine the role of the internal topological structure, as well as external factors such as product category and online reviews.

Given the interconnectedness of co-purchase relationships, it is essential to quantify co-purchases at the product-to-product level using a network approach. This study conducts systematic analyses to examine the formation of the co-purchase network. To do this, we first connect the products and generate a product co-purchase network if they were purchased by the same individual users, using a clickstream dataset from a real-world e-retailer. The network's topological structure is visualized and tested using a power-law approach, which reveals that the co-purchase network corresponds to a sparse small-world network, where most products have fewer co-purchase links and less sales volume, and a few products have a large number of co-purchase links and more volume. Our results show that the Pareto principle in product sales is validated from a network science perspective [4,49]. In a nutshell, a small proportion of products in the co-purchase network generates a large proportion of co-purchases and sales volume. In addition, different from the weak rich-get-richer effect in brand co-engagement behaviors [31], a Pearson correlation analysis suggests that co-purchase behavior has intrinsic concentration effects and a directed rich-get-richer concentration effect between the degree centrality and sales volume, where a product with a higher outgoing/incoming degree of centrality has a larger aggregated sales volume that originates from/targets to it. These findings suggest that the design of co-purchase recommendation systems can be made robust against random recommendation link failures. This conclusion is in line with previous research in the field [36,50]. If a co-purchase recommendation hyperlink fails or is inaccurate, it will not significantly impact the overall performance and accuracy of the system. This is because, with an understanding of the directed co-purchase relationships, it is easy to find a related product in a neighboring link by discovering the directed co-purchase patterns.

To further explore the inherent co-purchase pattern, we used ERGMs to examine the impact of the topological structure, product category, and online reviews on co-purchase. To the best of our knowledge, these are the first empirical results that identify the impact of product category and online reviews on the formation of co-purchase. Inferring cross-category relationship patterns among retail assortments enables effective product category promotions and recommendations [60,61]. The most relevant study to our research has a different objective than ours, as it focuses on the impact of recommendation systems on cross-category sales [4]. Their results show that the use of a recommendation system will reduce the diversity of the product categories that consumers view and purchase. Our results demonstrate that there are differential homophily effects between consumer cross-category purchase behaviors and co-purchase link formation. Another analysis of product cross-category purchases indicates that they significantly contribute to the formation of co-purchase relationships. This highlights the importance of identifying the appropriate product category for cross-selling or boosting sales.

Managing eWOM in the co-purchase network is a relatively new topic. Understanding of directed impact of eWOM on the formation of co-purchase relations is still scarce in recent studies [8,20,21]. Our results show that positive ratings and high review volume are key factors in predicting the formation of the co-purchase network. eWOM provides users with an eWOM persuasion mechanism for their co-purchase decisions. A highly positive rating of a product increases its sales, which helps consumers to build a perception of a product's quality [55]. Consumers' acceptable perception of high product quality can be deduced. When a user processes online review information and purchases a product with a highly positive rating, their demand for high-quality products might be met. This implies a lower probability of follow-up purchases for a relatively low-quality product. Furthermore, a higher inconsistency of positive ratings among products decreases the likelihood of co-purchase. This is similar to the self-reference effect in information decisions [56,57], where individuals have internalized information about the acceptable positive ratings related to themselves, and they react quicker to products with similar ratings. Additionally, the inconsistency effect of positive ratings has the strongest explanatory power. On the other hand, the higher the number of reviews, the more popular the product is, indicating that it has more disclosed information for consumers [52,62]. We further studied the effect of review volume on the dyadic product-to-product co-purchase research context. We found that a high review volume is positively associated with the log-odds of forming a co-purchase relation. This demonstrates the positive effect of product popularity on the formation of a co-purchase network.

5.2. Theoretical Implications

Our study makes three significant contributions to the literature on electronic networks and eWOM. Firstly, we extend the discourse on demand dependency [4–7] by presenting a data-driven explanatory framework that helps us to understand co-purchase decision making at the product-to-product level. This framework enables us to discover consumer co-purchase patterns from a network science perspective and to identify the factors that affect the formation of co-purchasing, rather than just examining the post hoc impact of co-purchase-network-based recommendation systems [9,10].

Secondly, we expand upon the existing knowledge of the statistical mechanics of consumer buying patterns by exploring the characteristics of product cross-category purchase behaviors [60,61]. Our study differs from prior research, which has mainly focused on the impact of recommendation systems on cross-category sales [4]. Our findings show the differential homophily effects of product category on the formation of co-purchasing, enabling us to uncover the hidden connections between products, which can provide insights for product placement and marketing strategies.

Lastly, this study also fills the research gap on the impact of eWOM on the formation of co-purchase relationships [8,20,21]. We include review volume and positive ratings as measures of the eWOM volume and valence. In particular, the negative impact of eWOM valence (i.e., positive ratings) suggests that, after a consumer purchases a product with a high positive rating, the probability of co-purchasing decreases. The inconsistency effect of positive ratings on the co-purchase links formation is significant and has the strongest explanatory power. Conversely, as the review volume of a product increases, indicating a greater disclosure of product information, the persuasive effect of eWOM is stronger, and the likelihood of co-purchasing increases. Our findings underscore the significance of managing eWOM in co-purchase behavior research.

5.3. Managerial Implications

From the perspective of managerial implications, this study aims to optimize cross-selling strategies and recommendation hyperlinks by discovering co-purchase patterns. To achieve this, the study develops an ERGM-based approach to understand the impact of online reviews and product categories on the formation of co-purchase relations in the form of a co-purchase network. The findings show the topological properties of the co-purchase network, the influence of online reviews on co-purchase decisions, and the appropriate product categories to place together in recommendation hyperlinks for cross-selling.

The study is applicable to different e-commerce platforms such as Amazon and JD.com. The findings of the consumers' cross-category purchase behavior can be used to improve the cross-selling efficiency and revenue in e-commerce. When it comes to the differential homophliy effects of product categories, there are seven categories of products in the network, wherein, product categories #1 and #5 are more likely to be co-purchased with the same category, while product categories #4 and #6 are highly associated with cross-category purchase. Moreover, the proposed ERGM can also infer intricate co-purchase patterns between different categories. These findings on consumers' cross-category purchase behavior can be applied within managerial strategies to improve the cross-selling efficiency and lift revenue of e-commerce.

Managing eWOM in the co-purchase network is a new topic with a potential impact for retailers. The consistency of the positive ratings between the products in co-purchase links is significant. According to the unique perception of the level of positive ratings that individual consumers can accept to buy a product, it is akin to a self-reference effect [56]. E-commerce should emphasize the exposure of products with close ratings in recommendation hyperlinks. This study also presents this important insight for e-commerce to place sponsor products for their users from the perspective of eWOM. According to the positive impact of a high review volume on consumers' co-purchase behaviors, online retailers should also encourage consumers to rate their products, to increase their popularity and information exposure. Additionally, e-commerce platforms should prominently feature high-volume review products (also indicated by frequent co-purchasing records in clickstream data) in recommendation hyperlinks on the product page or on the transaction page, in order to boost co-purchase sales.

6. Conclusions

In this paper, we present a framework with which to study the relationships between product co-purchasing, product categories, and eWOM. To do this, we used consumer clickstream data from a top Chinese online retailer to build a directed product co-purchase network. In this network, the products were represented as nodes and the directed links represented the co-purchase relationships. Using various statistical methods, including the power-law model and Pearson correlation test, we analyzed the topological structure of the co-purchase network. The extended ERGM approach was also employed to explore the impact of the topological structure, product category, and online reviews on the formation of co-purchase relationships. The results showed how eWOM (external mass opinions) and product category affect the association between products and the formation of directional co-purchase links.

This research endeavors to enhance cross-selling tactics and construct recommendation hyperlinks based on consumer behavior preferences and eWOM. Nevertheless, some limitations must be acknowledged. Due to the absence of data, the study was unable to assess the influence of web stimuli factors such as prices, promotions, and different types of recommendation systems on the formation of co-purchase networks. Furthermore, the perceived usefulness and helpfulness of online consumer reviews merit further investigation. Secondly, incorporating demographic information regarding the users would offer a more comprehensive understanding of the impact of collective behavior preferences on the network. Lastly, it would be an area of interest to examine the impact of higher-order network effects, such as triangles or more intricate relationships, on the co-purchase network in future research.

Funding: This research was funded by the Young Innovative Talents Research Project of Department of Education of Guangdong Province (2022WQNCX054).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aral, S.; Muchnik, L.; Sundararajan, A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21544–21549. [CrossRef] [PubMed]
2. McPherson, M.; Smith-Lovin, L.; Cook, J.M. Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **2001**, *27*, 415–444. [CrossRef]
3. Hill, S.; Provost, F.; Volinsky, C. Network-Based Marketing: Identifying Likely Adopters via Consumer Networks. *Stat. Sci.* **2006**, *21*, 256–276. [CrossRef]
4. Lee, D.; Hosanagar, K. How do recommender systems affect sales diversity? A cross-category investigation via randomized field experiment. *Inf. Syst. Res.* **2019**, *30*, 239–259. [CrossRef]

5. Celma, Ò.; Cano, P. From hits to niches? or how popular artists can bias music recommendation and discovery. In Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition, Las Vegas, NV, USA, 24–27 August 2008; pp. 1–8.
6. Fleder, D.; Hosanagar, K. Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. *Manag. Sci.* **2009**, *55*, 697–712. [CrossRef]
7. Lin, Z.; Goh, K.-Y.; Heng, C.-S. The demand effects of product recommendation networks: An empirical analysis of network diversity and stability. *MIS Q.* **2017**, *41*, 397–426. [CrossRef]
8. Kumar, A.; Hosanagar, K. Measuring the value of recommendation links on product demand. *Inf. Syst. Res.* **2019**, *30*, 819–838. [CrossRef]
9. Oestreicher-Singer, G.; Sundararajan, A. The visible hand? Demand effects of recommendation networks in electronic markets. *Manag. Sci.* **2012**, *58*, 1963–1981. [CrossRef]
10. Dhar, V.; Geva, T.; Oestreicher-Singer, G.; Sundararajan, A. Prediction in economic networks. *Inf. Syst. Res.* **2014**, *25*, 264–284. [CrossRef]
11. Huang, H.; Yao, L.; Chang, J.-S.; Tsai, C.-Y.; Kuo, R. Using product network analysis to optimize product-to-shelf assignment problems. *Appl. Sci.* **2019**, *9*, 1581. [CrossRef]
12. Kim, H.K.; Kim, J.K.; Chen, Q.Y. A product network analysis for extending the market basket analysis. *Expert Syst. Appl.* **2012**, *39*, 7403–7410. [CrossRef]
13. Mitra, A.; Ghosh, S.; Basuchowdhuri, P.; Shekhawat, M.K.; Saha, S.K. Recommendation system based on product purchase analysis. *Innov. Syst. Softw. Eng.* **2016**, *12*, 177–192. [CrossRef]
14. Monteserin, A.; Armentano, M. Influence-based approach to market basket analysis. *Inf. Syst.* **2018**, *78*, 214–224. [CrossRef]
15. Zhenghui, S.; Yun, H.; Sophia, F.J.; Mingxian, W.; Yan, F.; Noshir, C.; Wei, C. A Network-Based Approach to Modeling and Predicting Product Coconsideration Relations. *Complexity* **2018**, *2018*, 2753638.
16. Puka, R.; Jedrusik, S. A New Measure of Complementarity in Market Basket Data. *J. Theor. Appl. Electron. Commer. Res.* **2021**, *16*, 670–681. [CrossRef]
17. Ding, Z.; Hosoya, R.; Kamioka, T. Co-Purchase Analysis by Hierarchical Network Structure. PACIS 2018 Proceedings. 149. Yokohama, Japan. 2018. Available online: https://aisel.aisnet.org/pacis2018/149 (accessed on 26 August 2020).
18. Dellarocas, C. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Manag. Sci.* **2003**, *49*, 1407–1424. [CrossRef]
19. Filieri, R.; Hofacker, C.F.; Alguezaui, S. What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score. *Comput. Hum. Behav.* **2018**, *80*, 122–131. [CrossRef]
20. Stöckli, D.R.; Khobzi, H. Recommendation systems and convergence of online reviews: The type of product network matters! *Decis. Support Syst.* **2021**, *142*, 113475. [CrossRef]
21. Pan, X.; Hou, L.; Liu, K. The effect of product distance on the eWOM in recommendation network. *Electron. Commer. Res.* **2020**, *22*, 901–924. [CrossRef]
22. Lin, Z.; Wang, Q. E-commerce Product Networks, Word-of-mouth Convergence, and Product Sales. *J. Assoc. Inf. Syst.* **2018**, *19*, 23–39. [CrossRef]
23. Robins, G.; Pattison, P.; Kalish, Y.; Lusher, D. An introduction to exponential random graph (p *) models for social networks. *Soc. Netw.* **2007**, *29*, 173–191. [CrossRef]
24. Holland, P.W.; Leinhardt, S. An Exponential Family of Probability Distributions for Directed Graphs. *J. Am. Stat. Assoc.* **1981**, *76*, 33–50. [CrossRef]
25. Zhou, T.; Kuscsik, Z.; Liu, J.-G.; Medo, M.; Wakeling, J.R.; Zhang, Y.-C. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4511–4515. [CrossRef] [PubMed]
26. Al-Safi, J.; Kaleli, C. Item Genre-Based Users Similarity Measure for Recommender Systems. *Appl. Sci.* **2021**, *11*, 6108. [CrossRef]
27. Kafkas, K.; Perdahçı, Z.N.; Aydın, M.N. Discovering customer purchase patterns in product communities: An empirical study on co-purchase behavior in an online marketplace. *J. Theor. Appl. Electron. Commer. Res.* **2021**, *16*, 2965–2980. [CrossRef]
28. Kitsios, F.; Kamariotou, M.; Karanikolas, P.; Grigoroudis, E. Digital marketing platforms and customer satisfaction: Identifying eWOM using big data and text mining. *Appl. Sci.* **2021**, *11*, 8032. [CrossRef]
29. Liu, J.-G.; Zhou, T.; Che, H.-A.; Wang, B.-H.; Zhang, Y.-C. Effects of high-order correlations on personalized recommendations for bipartite networks. *Phys. A Stat. Mech. Its Appl.* **2010**, *389*, 881–886. [CrossRef]
30. Zhou, T.; Ren, J.; Medo, M.; Zhang, Y.-C. Bipartite network projection and personal recommendation. *Phys. Rev. E* **2007**, *76*, 046115. [CrossRef]
31. Xu, M.; Bhattacharyya, S. Inferring Brand Knowledge from Online Consumer Associative Brand Networks. In Proceedings of the AAAI Workshops, Vancouver, BC, Canada, 28 February–1 March 2022.
32. Gao, H.; Liu, H.; Yi, M. Inferring values of recommendation links: Analysis of co-purchase network based on ERGM and product involvement. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 15–17 January 2021; pp. 108–113.
33. Hunter, D.R.; Goodreau, S.M.; Handcock, M.S. ergm. userterms: A Template Package for Extending statnet. *J. Stat. Softw.* **2013**, *52*, i02. [CrossRef]

34. Hunter, D.R.; Handcock, M.S.; Butts, C.T.; Goodreau, S.M.; Morris, M. ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. *J. Stat. Softw.* **2008**, *24*, nihpa54860. [CrossRef]
35. Shmueli, G.; Koppius, O.R. Predictive analytics in information systems research. *MIS Q.* **2011**, *35*, 553–572. [CrossRef]
36. Gutiérrez-Moya, E.; Lozano, S.; Adenso-Díaz, B. Analysing the Structure of the Global Wheat Trade Network: An ERGM Approach. *Agronomy* **2020**, *10*, 1967. [CrossRef]
37. He, X.-J.; Dong, Y.-B.; Wu, Y.-Y.; Jiang, G.-R.; Zheng, Y. Factors affecting evolution of the interprovincial technology patent trade networks in China based on exponential random graph models. *Phys. A: Stat. Mech. Its Appl.* **2019**, *514*, 443–457. [CrossRef]
38. Deev, O.; Lyócsa, Š. Connectedness of financial institutions in Europe: A network approach across quantiles. *Phys. A: Stat. Mech. Its Appl.* **2020**, *550*, 124035. [CrossRef]
39. Sun, Y. How conversational ties are formed in an online community: A social network analysis of a tweet chat group. *Inf. Commun. Soc.* **2020**, *23*, 1463–1480. [CrossRef]
40. Bodin, Ö.; Nohrstedt, D. Formation and performance of collaborative disaster management networks: Evidence from a Swedish wildfire response. *Glob. Environ. Change* **2016**, *41*, 183–194. [CrossRef]
41. Kanrak, M.; Nguyen, H.O.; Du, Y. Maritime transport network analysis: A critical review of analytical methods and applications. *J. Int. Logist. Trade* **2019**, *17*, 113–122. [CrossRef]
42. Xie, J.; Bi, Y.; Sha, Z.; Wang, M.; Fu, Y.; Contractor, N.; Gong, L.; Chen, W. Data-Driven Dynamic Network Modeling for Analyzing the Evolution of Product Competitions. *J. Mech. Des.* **2020**, *142*, 031112. [CrossRef]
43. Kukar-Kinney, M.; Scheinbaum, A.C.; Orimoloye, L.O.; Carlson, J.R.; He, H. A model of online shopping cart abandonment: Evidence from e-tail clickstream data. *J. Acad. Mark. Sci.* **2022**, *50*, 961–980. [CrossRef]
44. Gao, H.; Liu, H.; Lin, W.; Chen, C. Graph-theoretic approach to detecting real-time intents within purchase conversion funnel using clickstream data. *Kybernetes* **2022**. *ahead-of-print*. [CrossRef]
45. Van den Poel, D.; Buckinx, W. Predicting online-purchasing behaviour. *Eur. J. Oper. Res.* **2005**, *166*, 557–575. [CrossRef]
46. Edmonds, P. Choosing the word most typical in context using a lexical co-occurrence network. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, Madrid, Spain, 7–12 July 1997; pp. 507–509.
47. Borgatti, S.P.; Everett, M.G.; Freeman, L.C. Ucinet for Windows: Software for social network analysis. *Harv. MA Anal. Technol.* **2002**, *6*, 12–15.
48. Xu, S.; Zhou, A. Hashtag homophily in twitter network: Examining a controversial cause-related marketing campaign. *Comput. Hum. Behav.* **2020**, *102*, 87–96. [CrossRef]
49. Brynjolfsson, E.; Hu, Y.; Simester, D. Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Manag. Sci.* **2011**, *57*, 1373–1386. [CrossRef]
50. Clauset, A.; Shalizi, C.R.; Newman, M.E. Power-law distributions in empirical data. *SIAM Rev.* **2009**, *51*, 661–703. [CrossRef]
51. Ou, C.X.J.; Chan, K.C.C. Developing a competitive edge in electronic markets via institutional and social based quality signaling mechanisms. *Inf. Manag.* **2014**, *51*, 532–540. [CrossRef]
52. Gu, B.; Park, J.; Konana, P. Research note—The impact of external word-of-mouth sources on retailer sales of high-involvement products. *Inf. Syst. Res.* **2012**, *23*, 182–196. [CrossRef]
53. Berger, J.; Sorensen, A.T.; Rasmussen, S.J. Positive effects of negative publicity: When negative reviews increase sales. *Mark. Sci.* **2010**, *29*, 815–827. [CrossRef]
54. Lee, J.; Park, D.-H.; Han, I. The effect of negative online consumer reviews on product attitude: An information processing view. *Electron. Commer. Res. Appl.* **2008**, *7*, 341–352. [CrossRef]
55. Forman, C.; Ghose, A.; Wiesenfeld, B. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Inf. Syst. Res.* **2008**, *19*, 291–313. [CrossRef]
56. Symons, C.S.; Johnson, B.T. The self-reference effect in memory: A meta-analysis. *Psychol. Bull.* **1997**, *121*, 371. [CrossRef] [PubMed]
57. Tam, K.Y.; Ho, S.Y. Understanding the impact of web personalization on user information processing and decision outcomes. *MIS Q.* **2006**, *30*, 865–890. [CrossRef]
58. Xu, D.; Ye, Q.; Hong, H.; Sun, F. Emotions for attention in online consumer reviews: The moderated mediating role of review helpfulness. *Ind. Manag. Data Syst.* **2022**, *122*, 729–751. [CrossRef]
59. Liu, J.; Mo, Z.; Gao, H. The effect of reviewer's review volumes on online consumer reviews' perceived usefulness: An event-related potential study. *Neurosci. Lett.* **2021**, *762*, 136139. [CrossRef] [PubMed]
60. Tanusondjaja, A.; Nenycz-Thiel, M.; Kennedy, R. Understanding shopper transaction data: How to identify cross-category purchasing patterns using the duplication coefficient. *Int. J. Mark. Res.* **2016**, *58*, 401–419. [CrossRef]

61. Mild, A.; Reutterer, T. An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data. *J. Retail. Consum. Serv.* **2003**, *10*, 123–133. [CrossRef]
62. Duan, W.; Gu, B.; Whinston, A.B. The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. *J. Retail.* **2008**, *84*, 233–242. [CrossRef]