JSDSS 日本ソーシャルデータサイエンス論文誌
第 9 巻　第 1 号（2025 年 3 月）

Comparison of Inter-Product Network Structure Between Physical Stores and E-Commerce Site

Jin Nakashima†1　Takashi Namatame†2　Kohei Otake†3

Abstract—With the rapid growth of electronic commerce (EC) and direct-to-consumer, retailers, in particular, are required to improve customer experience by promoting online-merges-with-offline (OMO) strategies. In this study, we constructed an inter-product network in brick-and-mortar stores and EC by defining the relationship between products based on the purchase-history data of consumer electronics stores and compared the structure of the network. This study aims to understand customers' purchasing behavior in both channels and discover knowledge useful for OMO measures. Specifically, we visualize the connection between products using the social network analysis method with the help of the research result that defines the similarity between products based on customer preferences from ID-POS data. The results suggest a difference in the network structure between products formed in physical and e-commerce stores.

Keywords—OMO，Omnichannel，Inter-Product Similarity，Social Network Analysis

1. Introduction

Since the online-merges-with-offline (OMO) concept was proposed in 2017, retailers have adopted a common practice to offer customers a fusion of online and offline options in their marketing activities [1]. The best example of an OMO measure is "buy online pick-up in store" (BOPIS). According to a study by Insider Intelligence, multichannel retailers such as Walmart and Home Depot are experiencing explosive sales growth, where the amount spent on BOPIS services was reported to be $72.46 billion in 2020, which is expected to increase to $140.96 billion in 2024 [2]. The survey results also highlight that companies that are not major players in the EC domain are ranked among the top companies regarding sales, indicating the importance of a seamless customer experience across multiple channels.

In Japan, OMO strategies are increasing in line with the growth of the electronic commerce (EC) market. Particularly, against the backdrop of the COVID-19 pandemic, the EC rate in the field of goods sales has been expanding rapidly, and the market scale of "direct to consumer" (D2C), which refers to a system for selling products directly to customers, is expected to reach 3 trillion yen by 2025 [3][4]. However, the monthly percentage of households using EC has been flat or declining, indicating that the number of new users is decelerating [5]. In the future, it will be necessary to increase customer engagement by providing value that transcends the boundary between online and offline, not only through simple purchasing but also by viewing the process from purchase through information search to consumption as a single experience.

In the academic field, few studies directly mention OMO, but through an understanding of customers' omnichannel shopping value (OCSV) and omnichannel intensity (OCI), the following studies were conducted. According to Huré et al., OCSV is a generic term for the cognitive, emotional, sensory, and social responses customers have to the channels offered by a company throughout the purchase process [6]. Lemon et al. highlighted that a comprehensive understanding of these experiences is necessary to understand the customer experience brought about by omnichannel [7]. Additionally, OCI measures the extent to which a company's multiple online and offline channels are integrated to provide a seamless shopping experience for customers [6]. By capturing customer experiences, we can evaluate the effectiveness of strategies based on channel fusion. Furthermore, Okutani presented the current status of omnichannel shopper research and the research issues required for understanding OCSV through a review of research areas in mobile app acceptance behavior, showrooming and webrooming, and BOPIS services contributing to understanding such consumer behavior [8]. Specifically, we derived three issues: understanding the role of mobile touchpoints as the nexus between online and offline, measuring OCI necessary for understanding OCSV, and evaluating the expected shopping experience of omnichannel shoppers who move across multiple channels.

Although previous studies have highlighted the importance of holistically understanding the shopping experiences of omnichannel shoppers, the experience indicators perceived by these customers have not been quantitatively captured. Therefore, this study attempts to understand the complex buying experience by quantitatively capturing customer preferences and visualizing their relationship structures. Particularly, we used social network analysis to express diverse customer preferences. Social network analysis is an analytical method that explores the relationships among the components of various objects. In general, data with a graph or network structure often demonstrate the relationships among objects represented as nodes. For example, the hyperlink structure between web pages and social networking services is widely known. By defining links based on such relationships among nodes, graph mining methods have been proposed, such as clustering based on a graph structure for community detection and identifying influential nodes based on PageRank scores [9][10].

†1 Graduate School of Tokai University
†2 Chuo University
†3 Sophia University (mail: k-otake@sophia.ac.jp)

In the marketing domain, purchasing and browsing relationships among products can be regarded as data representing connections among objects and are analyzed as graph structure data. Goto applied social network analysis to the problem of not being able to define relationships among nodes simply by defining relationships among webpages based on users' browsing histories [11]. Specifically, he reported that, in contrast to network graphs constructed from high-dimensional and sparse similarity among nodes, when network graphs are constructed in conjunction with dimensionality compression methods, nodes that are more closely related to each other may be connected, and effective clustering results may be obtained. Ito et al. defined the degree of similarity between products according to the number of customers who purchased common products based on ID-POS data and applied social network analysis [12]. Moreover, they extracted similar product groups through community detection and identified essential products using PageRank. These studies are valuable examples indicating that practical knowledge can be obtained from appropriate definitions of the relationships and similarities among objects.

Based on these research results, this study used social network analysis as part of an effort to quantitatively measure the comprehensive customer experience through channel fusion. We particularly defined the degree of similarity between products by using the ID-POS data of a Japanese retailer that operates electronic mass retailers in both brick-and-mortar and e-commerce channels and construct a product network between brick-and-mortar and e-commerce stores by using Ito et al.'s findings [12]. We also compared these structures regarding network indices to capture customer preferences and propose OMO measures that transcend the boundaries between channels.

2. Datasets

In this study, we analyzed purchase-history data obtained through purchases at Japanese electronics retail stores, which were both physical and e-commerce based. The stores sold products related to music, video, and game software as well as personal computers and peripherals. The data analysis period spanned one year, from January 1 to December 31, 2020. A summary of the data is presented in Table 1. In this study, three general merchandising stores handling various products were selected to analyze purchase-history data related to physical stores. These stores differ in the number of floors: Store A had one floor, Store B had three floors, and Store C had seven floors. The EC site data used were purchase-history data from an official general shopping site. The minimum classification category of the data was used for product classification in the inter-product network. The similarity among products was expressed as an edge list of Start-End combinations, and the similarity h_{ab} described below was assigned to each combination of products as a weight (Table 2). A node list representing each product category is created (Table 3). Using these edge and node lists, we constructed an inter-product network based on data from each of the three physical stores and e-commerce sites.

Table 1. Dataset Overview

Store Name	Number of Users	Number of Categories
Store A	12,704	521
Store B	8,443	570
Store C	24,474	681
EC	58,529	1,403

Table 2. Example of Edge List

Start	End	Weight
a	b	h_{ab}
b	c	h_{bc}
c	a	h_{ca}

Table 3. Exanple of Node List

Category ID	Category Name
a	Category A
b	Category B
c	Category C

3. Analytical Methods for Structuring Inter-Product Network

In this study, we constructed a network of products using the dataset described in Section II and compared their structures using network indices. Soft clustering by probabilistic latent semantic analysis (pLSA) was used to express customer preferences for the similarity between products based on ID–POS data.

3.1 Social Network Analysis

Social network analysis is an analytical method based on graph theory that explores the structure of the relationships among components in various objects, such as human relationships, distribution networks, and webpage links. In this study, we constructed a network (directed graph) of products by considering nodes as products and edges as similarities among products. The following network indicators were used to understand the network structure quantitatively:

- Total Degree: The total number of edges connected to a node is represented by Equation 1:

$$E = \sum_{i=1}^{V} d^{+}(v_i) + d^{-}(v_i) \tag{1}$$

  where V is the total number of nodes in the network and $d^{+}(v_i)$ and $d^{-}(v_i)$ are the outgoing and incoming orders of node v_i, respectively.

- Average Degree: The average number of connected edges per node can be obtained by dividing the total number of degrees by the number of nodes, as expressed in Equation 2:

$$average\ \ degree = \frac{E}{V} \tag{2}$$

- Density: The ratio of the actual number of edges to all possible edges in the graph is expressed by Equation 3.

$$density = \frac{E}{V(V-1)} \tag{3}$$

In the case of a directed graph, the number of edge combinations between two nodes is $V(V-1)$.

- Transitivity: A transitive relationship, that is, the degree to which a relationship such as "a friend of my friend is also a friend of my friend" is established, is represented by Equation 4 [13].

$$transitivity = \frac{3\delta}{t} \tag{4}$$

where δ is the total number of closed triangles in the graph, that is, the total number of unions at the three fully connected nodes, and t is the number of open triangle unions, including closed triangles, that are connected at two or more points. By multiplying the total number of closed triangles by 3, we considered the three patterns of open triangles contained within the closed triangles.

- Reciprocity: This is a measure of the proportion of mutually directed edge relationships in the entire graph; if the four patterns of directed edges between two vertices i and j are a, b, c, and d (Fig. 1), they are represented by Equation 5 [14].

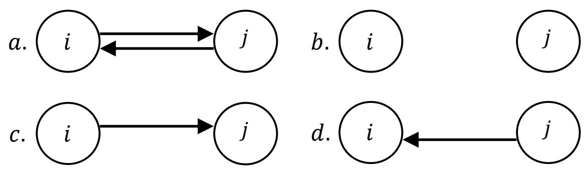

Fig. 1. The Pattern of Directed Edges Between Two Vertices $(i < j)$

$$reciprocity = \frac{a}{a+c+d} \tag{5}$$

3.2 Probabilistic Latent Semantic Analysis

In this study, pLSA [15]—a type of latent class model—was applied to define the similarity between products based on customer preferences. pLSA is a method for the dimensionality compression of data by explaining large high-dimensional data with several latent variables. In the target ID-POS data, soft clustering was performed based on the co-occurrence relationship between the purchased products (columns) and customers (rows). This makes it possible to model customer purchasing trends and product characteristics, while considering fuzzy similarity and diversity of customer preferences. Figure 2 presents a graphical model of pLSA.

Let $X = \{x_1, x_2, \cdots x_I\}$ be a set of I products, $Y = \{y_1, y_2, \cdots y_J\}$ be a set of J customers, and $Z = \{z_1, z_2, \cdots z_K\}$ be a set of K potential classes. In this case, the probability of event $P(x_i, y_j)$ in which customer y_j purchases product x_i is expressed in Equation 6.

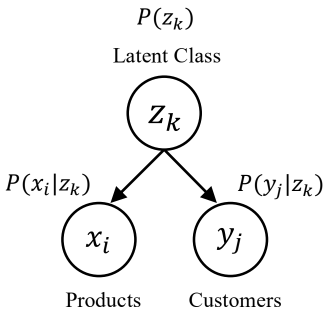

Fig. 2. Graphical Model of pLSA

$$P(x_i, y_j) = \sum_{k=1}^{K} P(z_k)P(x_i|z_k)P(y_j|z_k) \tag{6}$$

The parameters $P(z_k)$, $P(x_i|z_k)$, and $P(y_j|z_k)$ that maximized the logarithmic likelihood function LL in Equation 7 were estimated using the EM algorithm, which is an exploratory method [16]. $\delta(x_i, y_j)$ is an indicator function determined by whether customer y_j is purchasing product x_i.

$$LL = \sum_{i=1}^{I}\sum_{j=1}^{J} \delta(x_i, y_j) \log P(x_i, y_j) \tag{7}$$

4. Calculation of Inter-Product Similarity Based on Customer Preferences

In this study, the method proposed by Ito et al. was used to calculate the degree of similarity h_{ab} between products, as expressed in Equation 8 [12]. Initially, the similarity s_{ab} between products is calculated based on the relationships among customers purchasing common products. Moreover, we multiplied the weights w_{ab} based on the customer preferences obtained by pLSA and the probability of the inter-product transition t_{ab}. The product similarities h_{ab} obtained by this method were used to construct a product network.

$$h_{ab} = s_{ab} \times w_{ab} \times t_{ab} \tag{8}$$

where s_{ab} is the total number of customers who purchased both products a and product b in each set of analyzed data, and this is done for all combinations of the two products. By calculating the number of customers who purchased products a and product b, the network represents the connection between products that were likely to be purchased by a larger number of customers. Additionally, w_{ab} calculated using Equation (9) is a weight that considers customers' purchasing tendencies and product characteristics obtained using pLSA. When the probability that product i belongs to latent class k is θ_{ik}, the probability that product i belongs to each latent class is expressed as $\theta_i = (\theta_{i1}, \theta_{i2} \cdots, \theta_{iK})$ From this latent class distribution, the Jensen-Shannon divergence in Equation 10 can be used to obtain the distance between the distributions in the probability distribution θ_a of product a and θ_b of product b [17].

$$w_{ab} = 1 - D_{JS}(\theta_a || \theta_b) \tag{9}$$

$$D_{JS}(\theta_a || \theta_b) = \frac{1}{2}\left(D_{KL}(\theta_a || m) + D_{KL}(\theta_b || m)\right) \tag{10}$$

$$D_{KL}(\theta_a || \theta_b) = \sum_{k=1}^{K} \theta_{ak} \log \frac{\theta_{ak}}{\theta_{bk}} \tag{11}$$

$$\theta_{ik} = P(z_k | x_i) \tag{12}$$

$$m = \frac{1}{2}(\theta_a + \theta_b) \tag{13}$$

Because w_{ab} is treated as a distance, it has a symmetric relationship: $w_{ab} = w_{ba}$. The more similar the product affiliation probability distributions are, the stronger the connection. To consider the transition of purchasing tendencies, the probability of transition between products is defined as t_{ab}, where t_{ab} is the probability of transition between products. Specifically, we calculate the proportion of customers who purchased product b among those who purchased product a. The inter-product transition probabilities were added as weights, and an effective graph was constructed from the difference between the transition probability t_{ab} from product a to product b and the transition probability t_{ba} from product b to product a.

In this study, we employed the similarity h_{ab} between products as a means of attempting to quantitatively capture the cross-channel customer experience. This method is scalable to handle relatively large ID-POS data and is capable of obtaining universal and reliable findings for the analysis of a large number of customers and their purchase histories over a long period. Furthermore, the ability to handle graph-structured data without losing the diversity and heterogeneity of customer purchasing behaviors is considered an essential advantage for capturing complex customer experiences. Using these characteristics, we visualized ID-POS data as a network graph and quantitatively analyzed its structure to understand the customer experience in each channel. This is expected to have practical implications based on the interactions and synergies among the channels.

5. Structural Comparison of Inter-Product Networks

We visualized the connections among products by constructing an inter-product network using the similarity h_{ab} calculated based on the datasets of the three analyzed products (i.e., the three physical stores) and the e-commerce site. Furthermore, we quantitatively capture the structure of the network using the aforementioned network indices and compare customer preferences across channels. The graphopt layout algorithm [18], a dynamic model, was used to draw the network graph. The graph was pruned to 300 nodes from the viewpoint of its readability, leaving the edges with the highest similarity. The product network graphs for each store and e-commerce site are demonstrated in Figs. 3–6. The network indices are presented in Table 4.

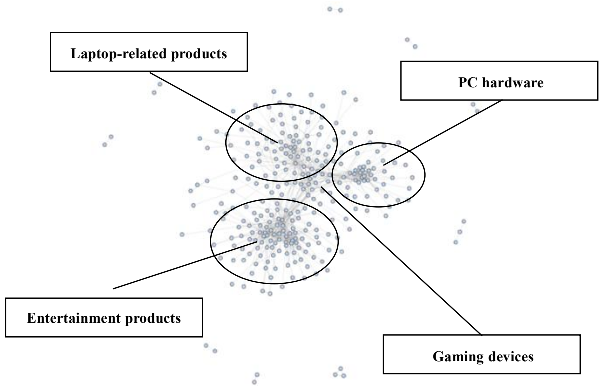

Fig. 3. Inter-Product Network of Store A

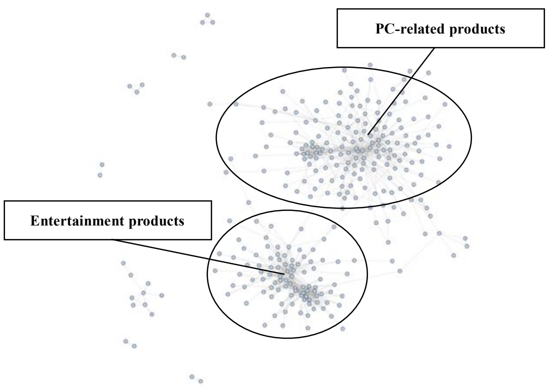

Fig. 4. Inter-Product Network of Store B

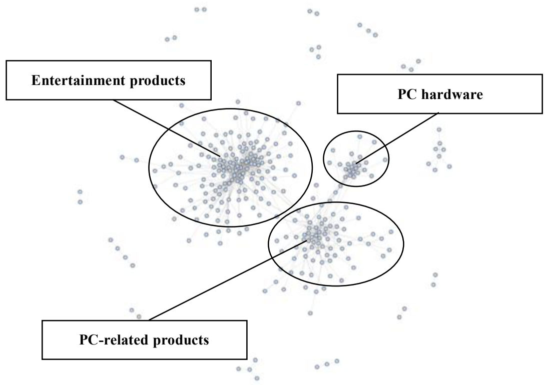

Fig. 5. Inter-Product Network of Store C

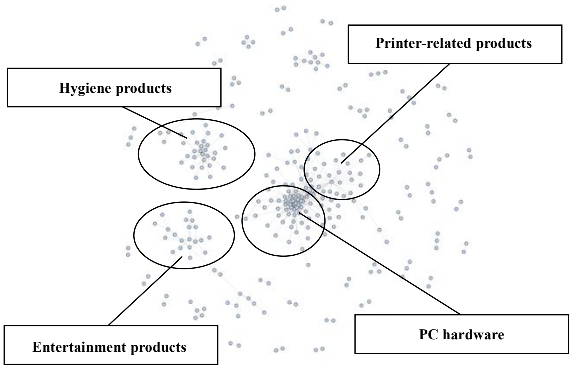

Fig. 6. Inter-Product Network of EC

Table 4. INDICES OF INTER-PRODUCT NETWORKS

Indices	Store A	Store B	Store C	EC
Connected Components	11	8	16	52
Total Degree	2,602	2,684	2,958	1,790
Average Degree	8.673	9.947	9.86	5.967
Density	0.015	0.015	0.016	0.001
Transistivity	0.291	0.337	0.4	0.46
Reciprocity	0.553	0.548	0.65	0.509

First, there are 11 connected components in the inter-product network of Store A, which is composed of multiple nodes with a small number of nodes and a single node with many nodes. Focusing on the largest central component, we observed that it is aggregated into three communities: laptop-related devices such as laptop computers and USB hubs; PC hardware such as motherboards, hard disks, and memories; and entertainment products such as game consoles, game software, and animation-related products. These communities were interconnected, which we inferred represents the purchasing behavior of customers in Store A that was structured on one floor and different product categories were often sold together.

Second, in the inter-product network of Store B, there were eight connected components: one consisting of a small number of nodes and one consisting of many nodes, similar to Store A. The largest connected component was the one that is visible from the overview. The largest connected component was divided into two communities, as shown in the overview, which were broadly classified into two categories: PC-related equipment, such as PC units and PC supplies, and entertainment products, such as games and animation-related products. Store B had three floors with used terminals on the first floor, PC-related products on the second floor, and entertainment products on the third floor, suggesting the formation of a polarized network structure.

Last, the inter-product network of Store C consisted of 16 connected components, and similar to Stores A and B, it had a structure in which a large connected component was attached to the center. Here, we noticed three major communities: entertainment products such as game consoles, game software, and animated BDs and DVDs; PC-related products such as PCs and PC supplies; and PC hardware such as hard disks and memories. Although similar to Store A at first glance, Store C had a structure in which the three communities were connected in series, rather than mutually. Store C had seven floors, which is a multilayered floor structure that may reflect the characteristics of purchasing behavior, such as weak or no connections between product categories across floors.

The inter-product network of an e-commerce site consisted of 52 connected components, which was significantly different from the structure of the inter-product network of a physical store described above. The large central component consisted of PC hardware, PC supplies, and printer-related products such as ink cartridges and printer paper. Other components were entertainment products such as games, animation, and plastic models, and hygiene products such as detergents and cosmetics. This is considered a tendency toward limited purchasing behavior. Therefore, it can be inferred that an EC site had a decentralized network structure that differed significantly from that of a physical store.

Comparing the network structures of physical stores and e-commerce sites based on the indices of each product network (Table 4), we noticed a significant difference, especially in order and density, and the e-commerce site was characterized by a slightly higher value of transitivity. As observed from the overview of the network graph, although there exist some differences in the way communities are formed owing to the physical-store structures, they basically form a coherent structure as one large connected component in the case of physical stores. By contrast, e-commerce-site structures are such that numerous connected components consisting of a small number of nodes are scattered, indicating that the connection between different products is weak compared to that of physical stores; that is, there is very little customer purchasing behavior across product categories. The relatively high transitivity value suggests network formation (triadic closure [19]), in which specific product groups are closely connected to each other and to related products of those that are already connected, which represents the aforementioned characteristics of the EC site channel.

Thus, a quantitative comparison of the network structure between products in a physical store and on an EC site that reflects customer preferences is expected to contribute to the proposal of measures for mutual customer traffic between channels from an OMO perspective. Specifically, the following measures can be considered:

- Enhancement of BOPIS Service: When a customer uses BOPIS in an EC or picks up a product at a store, similar products are suggested and arranged based on the product network between the physical store and EC to promote cross-selling. For example, when a customer at Store B orders gaming software, we suggest that anime-related CDs and DVDs are similar products in the EC. Furthermore, at the time of pickup, it is possible to implement dynamic measures such as placing a flash memory in the vicinity of a highly similar product in the store.

- Enhancement of EC Recommendation Function: In the inter-product network of a physical store, the recommendation of products that cannot be interpreted from the purchase history on the EC site can be utilized to create cross-selling opportunities. As a product recommendation widely connected to each community, it is considered effective to recommend a gaming device to a customer at Store A and a mouse or keyboard to a customer at Store C.

Particularly, in the case of e-commerce site recommendations, reflecting the preferences of stores frequently used by customers enables more personalized product recommendations and promotes additional purchases. Providing customers with a consistent purchasing experience and promoting integration among channels are crucial for effective OMO policy proposals.

6. Summary and Future Issues

In this study, we compared the structure of the inter-product network between physical stores and EC by defining the relationship between products based on customer preferences and purchase-history data of electronics retail stores. We defined the degree of similarity between products by multiplying the co-occurrence relations of purchased products, customer preferences using soft clustering, and transition probabilities between products and constructed an inter-product network. By quantitatively comparing the network graphs and network indices of each channel, we determined that communities influenced by the floor structure of a store are easily formed in a physical store, and that all of them have a certain degree of connection and are united. For e-commerce sites, the channel characteristics of the flow of customers through searches and recommendations resulted in a decentralized structure in which connections were easily completed among specific products. Furthermore, by focusing on the specific links between products, we understood that not only can the characteristics of each channel be understood, but useful suggestions can also be obtained for OMO measures based on customer preferences.

In the future, we plan to extract similar product groups and identify significant products by detecting communities and calculating centrality indices based on modularity. Another future task is to apply machine-learning methods to network graphs to understand the factors that form connections among products.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Numbers 21K13385 and 24H00370.

References

[1] beBit, Inc., "Beyond O2O, how was OMO born?" https://www.bebit.co.jp/blog/future/world-xd_07/, (author checked on 2024/6/28)

[2] Insider Intelligence, "Successful Strategies of the Top 10 E-commerce Companies in the U.S.," https://www.businessinsider.jp/post-217180, (author checked on 2024/6/28)

[3] Ministry of Economy, Trade and Industry, "2022 Market Research Report on E-Commerce," https://www.meti.go.jp/press/2022/08/20220812005/20220812005.html, (author checked on 2024/6/28)

[4] Internet Shop Managers Forum, "Digital D2C Market to Surpass 2 Trillion JPY by 2020 and Reach 3 Trillion JPY by 2025," https://00m.in/kmael, (author checked on 2024/6/28)

[5] Statistics Bureau, Ministry of Internal Affairs and Communications, "Household Consumption Survey Internet Shopping-Results for April 2024," https://www.stat.go.jp/data/joukyou/pdf/n_joukyo.pdf, (author checked on 2024/6/28)

[6] Huré, E., Picot-Coupey, K. and Ackermann, C. -L., "Understanding Omni-channel Shopping Value: A Mixed-Method Study," *Journal of Retailing and Consumer Services*, Vol. 39, pp. 314-330 (2017).

[7] Lemon, K. N. and Verhoef, P. C., "Understanding Customer Experience Throughout the Customer Journey," *Journal of Marketing*, Vol .80, No. 6, pp. 69-96 (2016).

[8] Okutani, T., "Understanding Omni-Channel Shopping Value from a Customer Perspective," *Japan Marketing Journal,* Vol. 40, No. 4, pp. 75-83 (2021) (in Japanese).

[9] Onizuka, M., "Application of Big Data with Graph Mining Technology and High-Speed Technology Initiatives," *Production and Technology*, Vol. 67, No. 2, pp. 4-10 (2015) (in Japanese).

[10] Girvan, M. and Newman, M. E. J., "Community Structure in Social and Biological Networks," *Proceedings of the National Academy of Science,* Vol. 99, No. 12, pp. 7821-7826 (2002).

[11] Goto, M., "Utilizing User Behavior History Data Network Analysis," *Communications of the Operations Research Society of Japan*, Vol. 64, No. 11, pp. 671-677 (2019) (in Japanese).

[12] Ito, H., Gendo, K. and Goto, M., "A study on Extraction of Important Items Focused on Customer Growth based on Network Analysis," *The 40th Symposium on Information Theory and its Applications* (2017) (in Japanese).

[13] Newman, M. E. J., Watts, D. J. and Strogatz, S. H., "Random Graph Models of Social Networks," *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 99, pp. 2566-2572 (2002).

[14] Suzuki, T., *Data Science in R 8: Network Analysis*, 2nd Edition, Kyoritsu Publishing, pp. 38-40 (2017) (in Japanese).

[15] Hofmann, T., "Probabilistic Latent Semantic Indexing," *Proceedings of the 22nd Annual ACM Conference on Research and Development in Information Retrieval*, pp. 50-57 (1999).

[16] Dempster, A. P., Laird, N. M. and Rubin, D. B., "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of Royal Statistical Society*, *Series B*, Vol. 39, NO. 1, pp. 1-38 (1977).

[17] Fuglede, B. and Topsoe, F., "Jensen-Shannon Divergence and Hilbert Space Embedding," *International Symposium on Information Theory*, pp. 31-39 (2004).

[18] The igraph core team, "The Graphopt Layout Algorithm," https://igraph.org/r/doc/layout_with_graphopt.html, (author checked on 2024/6/28)

[19] Granovetter, M. S., "The Strength of Weak Ties," *American Journal Sociology*, pp. 78, No.6, pp. 1350-1380 (2001).

