Inter-category Map: Building Cognition Network of General Customers through Big Data Mining

Gil-Young Song[1], Youngjoon Cheon[2], Kihwang Lee[3], Kyung Min Park[4], and Hae-Chang Rim[1]

[1]Department of Computer & Radio Communications, Korea University
Seoul, 136-701 – Republic of Korea
[e-mail: cielblau@korea.ac.kr, rim@nlp.korea.ac.kr]
[2]Research Center of Technological Management, Yonsei University
Seoul, 120-749 – Republic of Korea
[e-mail: taisaman@naver.com]
[3]The Mining Company, Daumsoft, Inc.
Seoul, 140-887 – Republic of Korea
[e-mail: leekh@daumsoft.com]
[4]School of Buisiness, Yonsei University
Seoul, 120-749 – Republic of Korea
[e-mail: kminpark@yonsei.ac.kr]
*Corresponding author: Kyung Min Park

Abstract

Social media is considered a valuable platform for gathering and analyzing the collective and subconscious opinions of people in Internet and mobile environments, where they express, explicitly and implicitly, their daily preferences for brands and products. Extracting and tracking the various attitudes and concerns that people express through social media could enable us to categorize brands and decipher individuals' cognitive decision-making structure in their choice of brands. We investigate the cognitive network structure of consumers by building an inter-category map through the mining of big data. In so doing, we create an improved online recommendation model. Building on economic sociology theory, we suggest a framework for revealing collective preference by analyzing the patterns of brand names that users frequently mention in the online public sphere. We expect that our study will be useful for those conducting theoretical research on digital marketing strategies and doing practical work on branding strategies.

Keywords: social media, big data mining, brand choice, inter-category map

1. Introduction

With the advent of the experience economy era, consumers' subjective opinions expressed in the online space have become an important force in driving social change. In particular, consumers' opinions displayed on social media are being used as a vital tool for grasping preferences and interests pertaining to products and even future buying decisions. Analyzing these opinions is essential when designing a branding portfolio. Moreover, marketing strategists can attempt to shape online public opinion in advance of the launch of a product, thus inducing positive consumer responses toward new products.

However, few theoretical or empirical studies have adequately addressed the question of how we should deal with complex consumer preferences. In economic sociology, it was pointed out that the boundaries of markets are subjectively formed by consensus between various audiences, i.e., market players and consumers, authorities. Therefore, the existence of subjective boundaries at the field level has been discussed [9,10,11]. These subjective boundaries are contrasted with traditional inter-market boundaries, which are determined by differences between industrial classification systems or products' technological characteristics. Products grouped into a single category from an objective perspective can be framed into a number of different identities from the diverse, subjective perspectives of consumers, which can vary across consumer types and time periods.

This study investigates the question of how seemingly unrelated product categories are perceived as similar by analyzing the opinions of online network service users. Based on this analysis, we present a network of consumers' preferences for, and awareness of, products; this is distinguished from an inter-actor network or an inter-organizational network in traditional sociological theory.

Naming this network of preferences an inter-category map, we show the possibility that consumer preferences expressed across multiple categories will form a market space and become a bundling solution for marketing. Our approach reveals the process of the subjective formation of market boundaries suggested by the theory of economic sociology [16,17]. Managers may regard this study as a useful framework for Internet-based collaborative marketing or alliance strategy.

Scholars recognize the practical impossibility of gathering data on consumers' cognitive decision-making structures. To overcome this obstacle, we utilized blog posts voluntarily written by numerous authors. This presents an opportunity to examine the real-time evolving market boundary constructed by consumers at a cognitive level [8,10].

This article is organized as follows. Section 2 outlines the previous work and theoretical background relevant to this study. In section 3, the data collection procedures and actual data analysis within our framework are described. Section 4 presents conclusions.

2. Background

2.1 Previous Work

Thus far, studies on consumer preferences observed in the online space have generally been concerned with the word-of–mouth (WoM) phenomenon [2,3]. These studies, being influenced by the traditional literature on diffusion, have dealt mainly with issues such as consumers' imitation behavior and social contagion. According to prevailing notions, if the number of users choosing a particular product increases or if a certain group of users is positioned in the center of a network, they would affect other users' purchases. Other studies have examined the corporate performance of user-generated content in the field of management information systems or organizational theory, advocating the benefits of co-creation [25,26]. This line of research, claiming that corporations and individuals develop creative solutions together to ensure profit and growth, was recently spotlighted in the quantitative marketing field [7,27]. To explain why individual users create new products on YouTube or in blogs and open-source communities, scholars suggest that the connection between existing user networks and the network of products plays a crucial role [11,23]. For example, a user's proactive behavior depends on the user's position (inside or outside) in a network [33,36].

The limitations of previous work on online word-of-mouth and users' opinions can be summarized as follows. First, previous works focused on applying current social network theory to the online environment, and most studies sought to explain the mutual influences between users in the online space rather than investigate the theoretical aspects of networks.

From our experiences, we know that online networks do not always provide us with a benefit. Recently, there have been some indications of information overload in the forms of commercial recommendations and messages coming from the networks surrounding consumers. Therefore, from the standpoint of a researcher such as Granovetter, who places more importance on obtaining values from individuals than from the structure itself, the preferences of online networks and individual actors can be regarded as over-socialized [12,15]. Many individual actors remain in a passive status, only being affected by the structure. Second, it is extremely difficult to analyze user preferences, as such preferences are revealed as a result of choosing specific products and not as intentions in most cases [6]. Directly observing purchase results and evaluations is not a trivial task. Alternative analysis frameworks or new perspectives for empirical studies are urgently needed.

2.2 Theoretical Background

2.2.1 Product Categorization by Consumers in the Online Space

Not many studies have dealt with the question of how product categorization infuences consumers in the online space. A number of marketing studies looked into the influences of brand or product categories on consumer preference. Studies based on cognitive science developed arguments regarding category changes with regard to brand extension. Delvecchio and Smith [39] suggested that users associate products having a brand extension with risk. They also claimed that a higher risk level induced a higher price-premium level due to brand extension. In general, marketing researchers have taken conservative stances with respect to brand extension [40].

Additionally, organizational ecology theorists, who have shown an interest in individual organizational forms since the late 1990s, argued that the identity of a product represents the identity of an organization, as the various internal architectures and investment strategies of an organization wield a certain level of influence during the design and deployment of a particular product [13, 41, 42]. Therefore, changing the product identity meant a change in the organizational routine, which is interpreted as a serious hazard that can even lead to organizational destruction from the point of view of ecology [17]. In the same context, the category, together with the network, was recognized as the most importance criterion for human decision-making. It is claimed that general consumers, being conservative, will not choose a product if it has unfixed and ambiguous specifications. It is also claimed that such ambiguity will eventually have negative effects on the financial performance of an organization. This view is very similar to the results of experiments and consumer surveys in the marketing-related literature. Recently, however, doubts have been raised concerning claims that only products with fixed categories can easily be chosen [44]. Pontikes [45] argued that the identity of a product can change according to the acceptance level of the audience, while Kim and Jensen [46] suggested that consumers' choices can be ch anged by an organization's way of dealing with consumers. For instance, when music consumers have a choice between experimental pieces and more traditional, well-balan ced pieces, the modern and experimental pieces typically garner the most positive rev iews from critics. In this case, we see the merit of an organization giving a product portfolio an ambiguous identity, as it can eventually secure good performance by expl oring its repertoire in the longer term.

Hannan, Polos, and Carroll [47] noted that decision-making related to the product category at the organizational level is relevant to competitive dynamics. Existing organizational ecology has evolved to incorporate arguments about the niche, i.e., the ecological space an individual organization occupies [49]. For example, generalists that expand their niche run a higher risk of extinction, whereas specialists concentrating on specific fields have lower levels of risk [50]. In the same way, products having wide categories have a high risk of being rejected by a conservative audience, but they can be very attractive to an audience sensitive to trends or when the environment is rapidly changing. Therefore, existing approaches with regard to categorical identity can have different effects depending on how we view them [45,47].

The online space can be viewed as a space in which the device of fuzziness is operating, as noted by Hsu [48]. Hsu also claimed that multiple category membership is a natural phenomenon, as blending and segregation simultaneously occur in the space where the organization exists, demolishing and generating old and new boundaries. Therefore, it is an important premise that online community users show preferences across multiple categories [51].

How, then, is an inter-category brand map related to different approaches regarding consumer selection? First, we need to consider the fact that consumer preference in an online environment is a rapidly changing and highly uncertain property [52]. In an online community, users easily change directions, following opinion leaders or seriously biased positions [11,32]. As a consequence, a product planner should take a different approach in the online space because an online network is a sphere in which negative signals can spread much more quickly as compared to the traditional sphere of public opinion [32,33]. In the online social space, an organization needs to embrace an audience with diverse properties. Given the social interaction of the audience, the existing boundaries of the product category and preference system inevitably become blurred.

2.2.2 Concept of Inter-market Boundary and Boundary Construction Based on Users' Cognitive Processing

At this point, it is necessary to consider the concept of the inter-market boundary before constructing the inter-category brand map. This concept was suggested by Burt [4,5] as a means of uncovering the destructive phenomenon of boundaries; Burt drew on social network theory to observe the resources and networks exchanged between various markets. In sociology, the market is interpreted as an entity constructed by agreement among multiple stakeholders without a pre-given industrial structure. Thus, the inter-market boundary can be changed through consumer conceptualization. To analyze consumers' various product choices from a social structural viewpoint, field-level analysis is required. In the early days, Burt posited the concept of a *network across the market* [1,4,5]. As the complexity of the economic ecosystem increases, there is a diversification of the parties concerned with corporate transactions. Consequently, connections between markets caused by trade and supply chain management came to be recognized as an important issue in relation to the network [10,14]. For instance, when a financial corporation requests that a system integration company build an information system, a connection between the financial market and the system integration service market is created. Additionally, there has been an attempt to understand correlations and similarities between markets by analyzing input-output tables [3,15]. In summary, most efforts to analyze the relationships between markets at the level of the entire social system have seemed to be of limited scope, focusing only on the inter-organizational network or on superficial relationships based on pre-determined industry classification data [18,20]. What if a connection between markets creates a network by grouping individual consumers' subjective preferences? In reality, consumers tend to create product bundles in various ways according to their own interests, and they purchase a bundle of products as a *combined solution*. Hence, connections between product categories and markets are created in consumers' cognitive decision-making models. Tilly stated that the task of categorizing specific objects and examining the networks among them is closely linked to the practice of social network analysis [38].

White, Godart, and Corona [54] described the behavior of an actor crossing the borders of multiple networks using the concept of *Netdom* (Network + domain), noting that the most favorable actor is a person who has domain-specific and various communication styles for multiple cliques. Similarly, Godart and White [53] theoretically proposed an actor who can show an ambiguous identity while *crossing or switching* multiple network domains. We observe that existing theoretical frameworks of sociology already have such a concept, whereby an actor can cross inter-market boundaries. This type of phenomenon can be recognized by identifying an actor who has interests in multiple domains instead of simply examining the relationships between actors.

However, to the best of our knowledge, no study has offered a concrete cognitive model from the viewpoint of social network theory. Thus far, most network studies have been unable directly to examine the relationship between cognitive decision-making and networks. For example, although friend networks at the personal level or reference networks at the researcher level have often been indirectly interpreted as an expression of *interest*, consumers' decision-making activities when categorizing products in various ways have not been analyzed [17,21]. Some studies show that, because monitoring between various people takes place in networks with *multiplexity*, *cohesiveness* in high-density networks or trust increases, while opportunistic behavior decreases [56]. This, however, is simply a network

property focusing on the relationships between human actors, not an element related to diverse user preferences.

Our study attempts to analyze the multiple categories phenomenon as experienced by users in their everyday lives and presented in the form of blog posts in the online space. People write about their own tastes and preferences in the online space. Studies of information systems call this social curation or regard it as a data consumption process that works through the mapping of information [55]. We conceptualize these elements as a major preference network forming users' social interests and suggest that this phenomenon can be abstracted as an inter-category brand map. Our study is expected to advance inter-market network research concerned with the exchange of information and resources that occurs in inter-corporate alliances or transaction networks.

We developed the foundations for an inter-category map by observing users' needs and opinions in the virtual space. By bridging the existing framework of the 'Network across the market' and 'Category and Identity in the social system', we shed new light on the complexity of the purchasing decisions of customers in the real world. Economic sociologists have tried to combine the network between actors and categorization of audience concepts; however, efforts to define the micro-cognitive side of decision making have been unsuccessful [4,15]. Therefore, considering consumers' decisions in the context of an inter-category map is theoretically significant.

First, from a theoretical standpoint, it is necessary to look at the problem of how to define a particular object through classification or categorization [16,31,59]. Researchers who study the impact of the organizational categorization on product performance in the market have emphasized the importance of clear categorization [10,22,31,62]. In other words, 'Authentic Identity'. Fitness between the traditional schema of the audiences and the firm or a product's identity is critical in this context; many researchers noted that a similar form of categorization between an actor and the mental model of an external audience leads to social adoption and the acquiring of cognitive legitimacy [14, 16]. These studies have also insisted that it is relatively difficult for a product or an organization with an ambiguous identity to guarantee its own legitimacy. On the other hand, some researchers assert the value of *ambiguity* in that most products and organizations have mixed identities, with those complexities creating opportunities to embrace diverse consumers [16,60]. According to White [31], *categorical fuzziness or syncopated complexity* is a factor common to modern industrial products and organizations. Thus, how products and organizations are mapped with other entities with different features is more important than how they are classified and categorized authentically. In particular, from a sociological perspective, the significance of markets and networks stems from the fact that they are constructed system rather than structurally given one; if this claim is accepted, then grasping the process of categorization by consumers (i.e., the audiences of products and organizations) has profound significance [28].

Methodologically, data collection and measuring methods that can capture individual consumers' awareness are urgently needed. At present, economic sociologists who study organizational forms or product forms and identities note that, by examining newspapers, magazines, and public reports by institutions, it is possible to gather data on how the surrounding environment recognizes organizations and products [14]. These scholars' main concern has been to determine how organizations and brands are evaluated by an external environment using content analysis or various coding methods [3,8,13]. This grew out of the tradition of sociology, which considered identity to be formed by external categorization as

Input Text (blog posting)	yojeum seumateupon-ilago hamyeon aipongwa gaelleogsiga 1,2wileul datugo issgo
Morphological Analysis	yojeum/NNIN2 seumateupon/NNIN2+i/CO+lago/ENCM ha/VBMA+myeon/ENCO2 aipon/NNIN2+gwa/PPCJ gaelleogsi/NNIN1+ga/PPCA1 1,2/NU+wi/NNDE2+leul/PPCA2 datu/VBMA+go/ENCO3 iss/AX+go/ENCO1+.../SC
Keyword Extraction	yojeum / seumateupon / aipon / gaelleogsi / 1,2wi / 1,2widatuda / datuda

Fig. 1. An example of language processing for blog posts

well as individual actors' voluntary expressions. However, the validity of the data collection process remains an issue, as it is also an indirect method. Moreover, most of the data are generated by critics, including journalists and experts, and not by consumers themselves. Therefore, it is necessary to discover a model in which an inter-market network is constructed through more practical and detailed consumer opinion analysis.

3. Data Analysis

3.1 Data Collection and Representation

To use consumers' awareness of individual brands directly, we use social media data. Social media, a type of Internet application developed based on what is termed Web 2.0, enables users to generate and share their own content. Because social media works in both the mobile and web environment in a highly interactive manner, it has brought many changes to the communication methods used by and among organizations, communities, and individuals [22].

As mentioned in the previous section, the argument that consumers provide crucial ideas and information to corporations in their product development process has often been raised in research regarding open innovation and co-creation. In this study, however, our intention is to reflect consumer preferences in their everyday lives by collecting textual information naturally written on social media.

There are many types of social media, and the boundaries between them are becoming more blurred [19, 23]. In general, the types of social media include wikis, blogs, micro-blogs, user-generated content services, and social network services. Here, we use blogs. We also considered using Twitter, which is attracting increasing numbers of users, but we decided to use only blogs because Twitter, with its 140-character length limitation and retweet functions, will show somewhat peculiar aspects. Moreover, tweets are not easily processable from a practical point of view due to their vast number. Most blog services in Korea are operated as free subscription-based services provided by large-scale portal sites. We only use blog posts produced in 2012 on one major blog service, Naver.

The first step of data preparation is to collect blog posts from the blog service site. We use a social big-data analysis platform called SOCIALmetrics™ (http://www.socialmetrics.co.kr), developed and operated by Daumsoft. The number of blog posts written in 2012 that we can access using SOCIALmetrics™ is well over 112,000,000. The number of authors who wrote at least one blog post in 2012 exceeds 3,900,000. These blog posts are the output of an intensive spam-removal process.

The second step is to convert the blog posts into a suitable format for analysis. It is essential to apply natural language processing and text mining technologies as a means of conversion, as blog postings are mostly composed of free text, a typical unstructured form of data [24, 25]. For our purpose, only morphological analysis and keyword extraction techniques are required. Fig. 1 depicts an example of the language processing procedure and outputs of each stage of the procedure.

After the language processing steps, the blog posts are represented using the vector space model, which is widely used in document retrieval systems [26].

	k_1	k_2	…	k_j
d_1	1	1	…	0
d_2	0	1	…	0
…	…	…	…	1
d_i	1	0	0	1

Fig. 2. Vector space model for blog posts

Every element of the document set $D = \{d_1, d_2, \cdots, d_i\}$ is represented by the vectors of the relevant occurrence indicator value for every element of the keyword set $K = \{k_1, k_2, \cdots, k_j\}$. The occurrence indicator value is determined by the following function:

$$O(d_i, k_j) = \begin{cases} 1, & k_j \text{ occurred in } d_i \\ 0, & k_j \text{ not occurred in } d_i \end{cases}$$

We only use $B = \{b_1, b_2, \cdots, b_j\}$, which is a subset of K and which consists of brand keywords such as *iPhone* and *Galaxy*, to represent blog posts.

We also convert the above into a *brand mention matrix*, which can be shown as follows:

	b_1	b_2	…	b_j
a_1	3	2	…	0
a_2	0	7	…	2
…	…	…	…	0
a_k	9	4	0	1

Fig. 3. A brand mention matrix

A brand mention matrix is a matrix of $A \times B$, where set $A = \{a_1, a_2, \ldots, a_k\}$ is a set of authors and B is a set of brands, as defined earlier. The value of each cell is determined by the function $M(a_k, b_j)$, which counts the number documents written by author a_k in which brand b_j is mentioned. As a result, we can effectively represent a particular brand by a vector

Brand 1	Brand 2	Jaccard	Brand 1	Brand 2	Jaccard	Brand 1	Brand 2	Jaccard
BMW	Benz	0.234611	Benz	Nike	0.101332	…	…	…
ToniMoly	Missha	0.208126	iPhone	Nike	0.078940	ToniMoly	Nikon	0.027482
iPhone	Galaxy	0.203850	Galaxy	Nike	0.069897	ToniMoly	Canon	0.026294
Nike	Adidas	0.193547	Galaxy	Canon	0.069813	Benz	ToniMoly	0.020225
Nikon	Canon	0.140633	iPhone	Canon	0.066186	BMW	ToniMoly	0.015931

Table 1. Pairwise Jaccard similarity coefficients of target brands (partial)

of the 'mention count' of each blog author. Often, we reduce mention counts to binary mention indicators.

Because the main purpose of this study is to show the usefulness of our framework, we selected 10 brands that belong to five product categories that frequently occur in blog posts. These are summarized in **Table 1**.

Category	Brand
SMART PHONES	iPhone, Galaxy
SPORTSWEAR	Nike, Adidas
IMPORTED CARS	BMW, Benz
CAMERAS	Nikon, Canon
ROAD SHOP BEAUTIES	Missha, ToniMoly

Table 1. Target categories and brands (Brands are written in English only for clarity.)

3.2 Analysis 1: Inter-category Similarity

As a first step, we attempt to measure the similarities between target categories. The underlying rationale is that categories C_i and C_j are similar if they share a relatively large number of authors who mentioned brands b_p, b_q, b_r, and b_s, where b_p and b_q belong to C_i and b_r and b_s belongs to C_j. Thus, we must measure the pairwise similarities for the 10 target brands before measuring the degree of inter-category similarity.

As described in the previous section, each brand is represented by a vector of the mention indicators of the authors. Thus, we can measure the similarities between brands by measuring the similarities between the vectors. There are many ways to measure vector similarities. We use the *Jaccard similarity coefficient* as calculated by the following simple formula [27]:

	IMPORTED CARS	SMART PHONES	SPORTSWEAR	CAMERAS	ROAD SHOP BEAUTIES
IMPORTED CARS		3.8742E-14	2.4350E-58	0.1834824	2.8115E-05
SMART PHONES	3.8742E-14		0.00054861	0.4978102	1.3730E-33
SPORTSWEAR	2.4350E-58	0.0005486		0.0013297	3.2617E-14
CAMERAS	0.18348249	0.4978102	0.00132975		0.60323282
ROAD SHOP BEAUTIES	2.8115E-05	1.3730E-33	3.2617E-14	0.60323282	

Table 2. The p-values of the χ^2 test inter-category preferences

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

In our case, brand pairwise Jaccard similarity coefficients are easily obtainable using the brand mention matrix filled with brand mention indicators. A partial result is shown in **Table 2**.

It is only natural that pairs belonging to the same category are highly similar. They tend to be mentioned together by numerous authors. On the other hand, there are cases in which brand pairs each belonging to different categories exhibit high degrees of similarity, such as (Benz, Nike), (iPhone, Nike), and (Galaxy, Canon). Hence, we note the possibility of the creation of an inter-market connection driven by consumers' subjective awareness, as mentioned earlier. Thus, we may also have to reevaluate the claims of economic sociologists who contend that clear category identification of products or organizations is important. It appears, rather, that whether a particular brand or product is associated with other categories could be the key to its success.

To obtain the inter-category similarities for our set of categories, we abstract the above result to the category level. **Fig. 4** shows this.
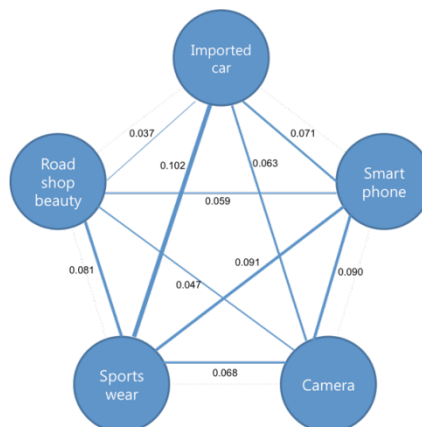

Fig. 4. Pairwise similarity for target categories

Category pairs that show high degrees of similarity are (IMPORTED CARS, SPORTSWEAR), (SMART PHONES, SPORTSWEAR), and (SMART PHONES, CAMERAS), in descending order. In contrast, category pairs that show low degrees of similarity are (IMPORTED CARS, ROAD SHOP BEAUTIES), (CAMERAS, ROAD SHOP BEAUTIES), and (SMART PHONES, ROAD SHOP BEAUTIES), in ascending order. The category ROAD SHOP BEAUTIES does not show a high degree of similarity with other brands but has a relatively high degree of similarity with the SPORTSWEAR category. This is an acceptable result, as beauty brands are normally associated with fashion brands and because sportswear brands and fashion brands share many common features.

3.3 Analysis 2: Inter-category Preferences

In this section, we observe the associations between categories when consumer preferences toward particular brands are considered.

Because we only apply a set of very simple language processing procedures, it is impossible fully to extract the exact brand preferences of consumers from blog posts. Consequently, we crudely define the brand preference as follows:

$$P(a_i, \langle b_p, b_q \rangle) = \begin{cases} b_p, & M(a_i, b_p) > M(a_i, b_q) \\ b_q, & M(a_i, b_q) > M(a_i, b_p) \end{cases}$$

where $M(a_i, b_p) > 0, M(a_i, b_q) > 0$.

The above definition of brand preference may not reflect actual consumer preferences. However, because we use brand preferences gathered from a large amount of data, it is still a useful and viable alternative to more accurate consumer preferences.

Using the above preference measure, we were able to observe the associations between pairs of categories. Pairwise associations are tested using the χ^2 test, as shown in **Table 3**. Some of the raw statistics are as follows (expected frequencies are shown in parentheses):

(SMART PHONES, IMPORTED CARS)

	BMW	Benz	Sum
iPhone	3725 (3528)	3645 (3842)	7370
Galaxy	1877 (2074)	2457 (2260)	4334
Sum	5602	6102	11704

p-value: 3.8742E-14
χ^2: 2.358E-27

(SMART PHONES, SPORTSWEAR)

	Nike	Adidas	Sum
iPhone	7394 (7928)	2858 (2954)	10252
Galaxy	4098 (4194)	1793 (1697)	5891
Sum	11492	4651	16143

p-value: 0.0005486

χ^2: 4.7278E-07

(SMART PHONES, CAMERAS)

	Nikon	Canon	Sum
iPhone	3748 (3727)	5849 (5870)	9597
Galaxy	2745 (2766)	4378 (4357)	7123
Sum	6493	10227	16720

p-value: 0.4978102

χ^2: 0.4503

(SMART PHONES, ROAD SHOP BEAUTIES)

	ToniMoly	Missha	Sum
iPhone	2041 (2339)	3876 (3578)	5917
Galaxy	2074 (1776)	2417 (2715)	4491
Sum	4115	6293	10408

p-value: 1.370E-33

χ^2: 2.9612E-66

The results are quite similar to those in 3.2. The ROAD SHOP BEAUTIES category is least associated with other brands, while the SMART PHONES and IMPORTED CARS categories are closely associated with other categories, apart from the ROAD SHOP BEAUTIES category.

Looking into preferences at the brand level reveals several interesting facts. For example, if some authors prefer a BMW over a Mercedes Benz, they tend to prefer an iPhone over a Samsung Galaxy. Moreover, if some authors prefer an iPhone over a Galaxy, they tend to prefer Nike over Adidas shoes.

	SMART PHONES	Prefer iPhone	Prefer Galaxy	Diff.
IMPORTED CARS	0.071	0.064	0.046	**0.018**
ROAD SHOP BEAUTIES	0.059	0.048	0.045	0.003

CAMERAS	0.090	0.071	0.064	0.007
SPORTS WEAR	0.091	0.081	0.055	**0.025**

Table 4. Differences in similarities between SMART PHONE preferences and other categories

The SMART PHONE category is by far the most frequently mentioned category in terms of preferences. According to **Table 4**, the brand preference differences for the SMART PHONE category are pronounced when they are compared to IMPORTED CARS and SPORTSWEAR categories. The exact meaning of this phenomenon is yet to be determined. At the very minimum, we note that authors who express preferences in the IMPORTED CARS and SPORTSWEAR categories show clear and sharp preferences in the SMART PHONE category.

If we go further, we expect that we can make a substitute or a proxy category chain model where one category can be substituted for another category for consumer preference inferences if the preference data for the object category are missing or cannot be measured.

4. Conclusion

4.1 Summary

Is it possible to predict a preference toward a particular brand from the brand preferences of other categories? In this article, we examined the relationships between the brand preferences of multiple categories framed by consumers based on the concept of a *market as a constructed entity* discussed in social network theory and economic sociology in the past.

Previous studies of networks across markets have suggested, by analyzing inter-corporation transaction histories or input-output tables, that an inter-organized network creates a connected market [5,31,61]. These studies, however, could not incorporate consumer awareness, which serves as one of the main actors in an inter-organized market. Due to the difficulties of collecting reliable data, the majority of previous studies treated consumers' roles only as a dependent or interaction variable derived from surrogate measures such as sales records or evaluations by critics [9,13,17].

Our approach used consumer data directly extracted from social media data in which consumers explicitly mentioned brands. Consequently, we were able to build a formal framework that can reveal the collective extent of awareness latent in consumers' cognitive decision-making structures. We believe that investigating the extent of awareness of consumers with an inter-category map can promote more abundant and lively discussions. Such an approach complements traditional research on categorization, which primarily focuses on market segments [23,24,29,58]. Our framework is also expected to offer the opportunity to observe the process whereby consumer awareness disperses dynamically from one category to other categories when incorporating longitudinal elements into the framework. Therefore, we may expect new discourse in the field of social science, with empirical studies and product market modeling making use of the methodology laid out in this article.

From a marketing strategy point of view, our framework can aid manufacturers in their inter-category marketing efforts to engage the consumer product categorization process that

is created in social media channels. Specifically, our methodology would be useful to induce consumers to prefer certain brands in various categories by offering targeted and customized recommendation services.

4.2. Practical Implications

In this article, we have shown that the categories cognitively formed by consumers play important roles in establishing market boundaries. It would be possible to use this phenomenon for strategic planning as it relates to product recommendations and bundling and for extrapolating consumer attitudes toward various products.

From the perspective of digital marketing, these results imply that marketers can apply an inter-category map in the process of intentionally building word-of-mouth and diffusing information on specific products [7,32]. Using the inter-category map, it would also be possible to influence consumers' perspectives by indirectly shaping public opinion.

On the other hand, we can also consider an inter-organizational alliance strategy beyond the product market boundary [15,18,57,41]. Most existing alliance strategies are assumed to create an intra-industry network at the local market level or to procure resources from an external market to produce certain products or services. However, once a cognitive map of brands categorized by customers is obtained using the inter-category map, heterogeneous brands belonging to different industries would be good candidates for alliance partners with which to construct networks for reciprocal alliances.

In general, the performance of existing alliances can only be measured indirectly based on the construct of organizational financial performance. However, the inter-category map enables us comprehensively to observe the complex attitudes of customers toward bundles of various resources. Thus, we expect that the inter-category map can be used as a crucial criterion for evaluating the performances of alliance marketing operations.

Lastly, we infer that the inter-category map has its own role in managing business portfolios when merging and acquiring brands. Therefore, we suggest that the concept of the inter-category map is related to the literature on real options or diversification.

4.3 Future Directions

This study does have limitations, which present opportunities for future research. First, the usefulness of the inter-category map was shown only at the descriptive statistics level. Therefore, it lacks the more rigorous implications that can be gained when using precise statistical methods. It would be desirable to conduct a consumer survey as a comparative and supporting study. Second, the current study only covered five target categories. We need to extend our coverage to other industries such as the service and manufacturing industries. Third, we can apply more sophisticated language processing and text mining technologies such as synonym processing and associative term analysis. By doing so, we will be able to extract more complete brand mention data from social media data. It will also be possible to segment consumer preferences or interests according to relevant features such as design, price, and function. Fourth, as stated earlier, we may be able to apply the current framework to Twitter data. This may uncover media-specific characteristics of consumer preferences.

References

[1] Abolafia, M.Y. 1984. "Structured Anarchy: Formal Organization in the commodities futures markets", Adler, P. and Adler, P. (eds). *The Social dynamics of Financial Markets Greenwich*, CT: JAI Press, pp. 129-152.

[2] Aral, S. Walker, D. 2011, "Creating social contagion through viral product design: A randomized trial of peer influence in networks", *Management Science*, vol. 57, no. 9, pp. 1623-1639. Article (CrossRef Link)

[3] Bradach, J.L. and Eccles, R.G. 1989. "Price, Authority, and Trust: From ideal types to plural forms", *Annual Review of Sociology*, vol. 15, pp. 97-118. Article (CrossRef Link)

[4] Burt, R, 1992. *Structural holes: The social structure of competition*, Cambridge, MA: Harvard University Press.

[5] Burt, R. 2000. "The network structure of social capital", *Research in Organizational Behavior*, vol. 22, pp. 345-423, New York: Elsvier, JAI Press. Article (CrossRef Link)

[6] Coleman, J.S. 1984. "Introducing social structure into economic analysis", *American Economic Review*, vol. 74, Papers and proceedings, pp. 84-88. Article (CrossRef Link)

[7] Dellarocas, C. 2003. "The digitization of Word of Mouth: Promise and challenges of online reputation", *Management Science*, vol. 49, no. 10, pp. 1407-1424. Article (CrossRef Link)

[8] Dimaggio, P.J. 1991. "Constructing an organizational field as a professional project: U.S. Art museums, 1920-1940." W.W. Powell and P.J. Dimaggio (eds.), *The New Institutionalism in organizational analysis*: 267-292. Chicago: University of Chicago press.

[9] Faulkner, R.R. 1983. *Music on Demand: Composers and Careers in the Hollywood film industry*, New Brunswick, NJ: Transaction Books.

[10] Fligstein, N. 2002. "Social Skill and the theory of fields", *Sociological Theory*, vol. 19, no. 2, pp. 105-125. Article (CrossRef Link)

[11] Goldenberg, J. Han, S. Lehmann, D. and Hong, J. 2009. "The Role of Hubs in the Adoption Process", *Journal of Marketing*, vol. 73, no. 2, pp. 1-13. Article (CrossRef Link)

[12] Granovettor, M. 1985. "Economic Action and Social structure: The problem of embeddedness", *American Journal of Sociology*, vol. 91, no. 3, pp. 481-510. Article (CrossRef Link)

[13] Glynn, M.A. and Lounsbury, M. 2005. "From the critics' corner: Logic blending, discursive change and authenticity in a cultural production system", *Journal of Management Studies*, vol. 42, no. 5, pp. 1031-1055. Article (CrossRef Link)

[14] Hsu, G. and Hannan, M. 2005, "Identity, Genre and Organizational Form", *Organization Science*, vol. 16, no. 5, pp. 474-490. Article (CrossRef Link)

[15] Hirsch, P.M. 1972. "Processing fads and fashions: An Organizational set analysis of cultural industry systems", *American Journal of Sociology*, vol. 77, no. 4, pp. 639-659. Article (CrossRef Link)

[16] Lounsbury, M. 2001. "Institutional sources of practice variation: Staffing college and university recycling", *Administrative Science Quarterly*, vol. 46, no. 1, pp. 29-56. Article (CrossRef Link)

[17] Lounsbury, M. and Glynn, M. 2001. "Cultural entrepreneurship: Stories, legitimacy, and the acquisition of resources", *Strategic Management Journal*, vol. 22, no. 6-7, pp. 545-564. Article (CrossRef Link)

[18] Obstfeld, D. 2005, "Social networks, the tertius iungens orientation, and involvement in innovation", *Administrative Science Quarterly*, vol. 50, no. 1, pp. 100-130. Article (CrossRef Link)

[19] Pallis, D. Zeinalipour, M. D. Dikaiakos, 2011, "Online Social Networks: Status and Trends", *Web Data Management Trails*, editors L. Jain and A. Vakali, Springer.

[20] Podolny, J. 1993. "A status based model of market competition", *American Journal of Sociology*, vol. 98, no. 4, pp. 829-872. Article (CrossRef Link)

[21] Polos, L. and Hannan, M.T. 2002, "Reasoning with partial knowledge", *Sociological Methodology*, vol. 32, no. 1, pp. 133-181. Article (CrossRef Link)

[22] Rao, H., Monin, P. and Durand, R. 2005. "Border Crossing: Bricolage and the erosion of categorical boundaries in French gastronomy", *American Sociological Review*, vol. 70, no. 6, pp.

968-991. Article (CrossRef Link)

[23] Soh, C., Markus, M.L., and Goh, K.H. 2006. "Electronic Marketplace and Price transparency: Strategy, Information Technology and Success", *MIS Quarterly*, vol. 30, no.3, pp. 705-724. Article (CrossRef Link)

[24] Simmel, G. 1950. "The Triad", Kurt H. Wolff (eds), *The Sociology of Georg Simmel*, New York: Free Press, pp. 45-169.

[25] Von Hippel, E, 1994. "Sticky Information and Locus of problem solving: Implication for Innovation", *Management Science*, vol. 40, no. 4, pp. 429-439. Article (CrossRef Link)

[26] Von Hippel, E. 1994. "Lead Users, A source of novel product concepts", *Management Science*, vol. 32, no. 7, pp. 791-805. Article (CrossRef Link)

[27] Von Krogh, G. and Von Hippel, E. 2006. "The promise of Research on Open Source Software", *Management Science*, vol. 52, no. 7, pp. 975-983. Article (CrossRef Link)

[28] Weber, M. 1947. "Sociological Categories of Economic Action", Talcott Parsons (eds.), *The theory of social and economic organization*, NY: Free press, pp. 158-323.

[29] Wellman, B. 1988. "Structural Analysis: From method and Metaphor to Theory and Substance", Barry Wellman and Berkowitz (eds). *Social Structures: A network approach*, NY, Cambridge University Press, pp. 19-61

[30] Weick, K.E. 1995. "Sensemaking in organizations", *Thousand Oaks*, CA: Sage Publication.

[31] White, H.C. 1993. "Markets in Production Networks", Richard Swedberg (eds.), *Explorations in Economic Sociology*, New York: Russell Sage Foundation, pp. 161-175.

[32] Kaplan A. M., Haenlein, M. 2010. "Users of the world, unite! The challenges and opportunities of social media", *Business Horizons*, vol. 53, no. 1, p. 59-68. Article (CrossRef Link)

[33] H. Kietzmann, Jan & Kristopher Hermkens. 2011. "Social media? Get serious! Understanding the functional building blocks of social media", *Business Horizons*, vol. 54, no. 3, pp. 241–251. Article (CrossRef Link)

[34] Feldman, R., and Sanger, J. 2006. *The Text Mining Handbook*. New York: Cambridge University Press.

[35] Kao, A., and Poteet, S. (eds). 2006, *Natural Language Processing and Text Mining*. Springer.

[36] G. Salton, A. Wong, C. S. Yang. 1975. "A vector space model for automatic indexing", *Communications of the ACM*, vol.18 no. 11, pp. 613-620. Article (CrossRef Link)

[37] Tan, Pang-Ning, Steinbach, Michael, and Kumar, Vipin. 2005. *Introduction to data mining*, Addison-Wesley.

[38] Tilly, C. 1998, *Durable Inequality*, University of California press, Berkeley and Los Angeles, California.

[39] DelVecchio, D., & Smith, D. C. 2005. "Brand-extension price premiums: the effects of perceived fit and extension product category risk", *Journal of the Academy of Marketing Science*, vol. 33, no. 2, pp. 184-196. Article (CrossRef Link)

[40] Roth, M. S., & Romeo, J. B. 1992. "Matching Product Category and Country Image Perceptions: A Framework for Managing Country-Of-Origin Effects", *Journal of International Business Studies*, vol. 23, pp. 477-497. Article (CrossRef Link)

[41] Negro, G., Hannan, M. T., & Rao, H. 2011. "Category reinterpretation and defection: Modernism and tradition in Italian winemaking", *Organization Science*, vol. 22, no. 6, pp. 1449-1463. Article (CrossRef Link)

[42] Koçak, Ö., Hannan, M. T., & Hsu, G. 2009. Audience structure and category systems in markets. In *Organizational Ecology Workshop*, Verona.

[43] Burt, R. S., & Talmud, I. 1993. "Market niche". *Social Networks*, vol. 15, no. 2, pp. 133-149. Article (CrossRef Link)

[44] Phillips, D. J., & Zuckerman, E. W. 2001. "Middle‐Status Conformity: Theoretical Restatement and Empirical Demonstration in Two Markets". *American Journal of Sociology*, vol. 107, no. 2, pp. 379-429. Article (CrossRef Link)

[45] Pontikes, E. G. 2012. "Two Sides of the Same Coin How Ambiguous Classification Affects Multiple Audiences' Evaluations". *Administrative Science Quarterly*, vol 57, no.1, pp. 81-118. Article (CrossRef Link)

[46] Kim, B. & Jensen, M. 2011. "How Product order affects Market identity: Repertoire ordering in the U.S. Opera Market", *Administrative Science Quarterly*, vol. 56, no. 2, pp. 238-256. Article (CrossRef Link)

[47] Hannan, M. T., Pólos, L., & Carroll, G. R. 2007. *Logics of organization theory: Audiences, codes, and ecologies*. Princeton University Press.

[48] G. Hsu, 2006. "Evaluative schemas and the attention of critics in the US film industry". *Industrial and Corporate Change*, vol. 15, no. 3, pp. 467-496. Article (CrossRef Link)

[49] Han, J. 2004. "Network Across Markets and Ecology of Organization: Dynamics of Manufacturing Firms in Korea, 1981-1999", *Korean Journal of Sociology*, vol. 38, no. 4, pp. 187-214. Article (CrossRef Link)

[50] Dobrev, S. D., Kim, T, & Carroll, G. 2003. "Shifting Gears, Shifting Niches: Organizational Inertia and Change in the Evolution of the US Automobile Industry, 1885–1981". *Organization Science*, vol. 14, no. 3, pp. 264-282. Article (CrossRef Link)

[51] Huang, E., Davison, K., Shreve, S., Davis, T., Bettendorf, E., & Nair, A. 2006. "Facing the Challenges of Convergence Media Professionals' Concerns of Working Across Media Platforms". *Convergence*, vol. 12, no.1, pp. 83-98. Article (CrossRef Link)

[52] Van Alstyne, M., & Brynjolfsson, E. 2005. "Global village or cyber-balkans? Modeling and measuring the integration of electronic communities". *Management Science*, vol. 51, no. 6, pp. 851-868. Article (CrossRef Link)

[53] Godart, F. C., & White, H. C. 2010. "Switchings under uncertainty: The coming and becoming of meanings". *Poetics*, vol. 38, no. 6, pp. 567-586. Article (CrossRef Link)

[54] White, H. C., Godart, F. C., & Corona, V. P. 2007. "Mobilizing identities: Uncertainty and control in strategy". *Theory, culture & society*, vol. 24, no. 7-8, pp. 181-202. Article (CrossRef Link)

[55] Duh, K., Hirao, T., Kimura, A., Ishiguro, K., Iwata, T., & Yeung, C. M. A. 2012. May. "Creating Stories: Social Curation of Twitter Messages". In ICWSM.

[56] Kim, T. Y., Oh, H., & Swaminathan, A. 2006. "Framing interorganizational network change: A network inertia perspective". *Academy of Management Review*, vol. 31, no. 3, pp. 704-720. Article (CrossRef Link)

[57] George, G. & Bock, A. J. 2011. "The Business Model in practice and its implication for entreprenuership research", *Entrepreneurship theory and practice*, vol. 35, no. 1, pp. 83-111. Article (CrossRef Link)

[58] Park, H, Friston, K. 2013. "Structural and functional brain networks: from connections to cognition". *Science*, vol. 342, no. 6158. Article (CrossRef Link)

[59] Bhaskarabhatia, A. & Klepper, S. 2014. "Latent submarket dynamics and industry evolution: Lessons from the US laser industry", *Industrial and Corporate Change*, vol. 23, no. 1. Article (CrossRef Link)

[60] Barnett, W. Feng, M. & Luo, X. 2013. "Social identity, market memory and first mover advantage", *Industrial and corporate change*, vol. 22, No. 3, pp. 585-615. Article (CrossRef Link)

[61] Negro, G. Hsu, G. & Kocak, O. 2010. "Research on categories in the sociology of organizations", *Research in the Sociology of Organizations*, vol. 31, pp. 3-35. Article (CrossRef Link)

[62] Carroll, G. Feng, M. Le Mens, G. Mckendrick, D. 2010. "Organizational evolution with fuzzy technological formats: Tape drive producers in the world market, 1951-1998", *Research in Sociology of Organizations*, vol. 31, pp. 203-233. Article (CrossRef Link)

Gilyoung Song received B.S. degree in computer science from Korea University, Seoul, Korea in 1992. After receiving M.S. degree from the same university in 1994, Gil-Young worked at Korea Trade Network in Seoul as a researcher and consultant carrying out various international trade automation and e-government projects. He has played a vital role in nurturing Daumsoft, Inc. as a chief strategy officer and now as a vice president since joining the company in 2001. Daumsoft, Inc. is a company pioneering the science of translating on-line consumer opinions into actionable intelligence for real world business practice. Gil-Young is also actively involved in a number of academic communities trying to bridge the gap between academics and industry. He received Ph.D. in computer science from Korea University in 2014.
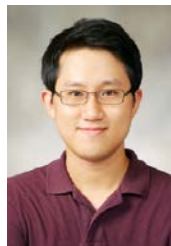
Youngjoon Cheon is a Ph.D candidate at the graduate school of Management of Technology, Yonsei University, Seoul, Korea. He received M.S. degree in information systems engineering from Yonsei University. His research interests include technological Strategy, Economic Sociology and Artistic creation and collaborative network. He is also a senior researcher of Technological Management research center in Yonsei University. His research has been appeared in Asian Social Science, Personal and Ubiquitous Computing, Korean Management Research.

Kihwang Lee received B.A. and M.A. degrees in Korean linguistics from Yonsei Universty, Seoul, Korea in 1991 and 1993, respectively. After receiving Ph.D. in informatics majoring computational linguistics from the University of Edinburgh, U.K. in 2005, Kihwang worked as a senior researcher, and later research professor in Yonsei University. From 2011, he is leading a team responsible for developing and operating social big data mining system at Daumsoft, Inc., a company specialized in trend spotting and insight finding through text mining and big data processing. His research interests include computational data-intensive linguistics, digital humanities, and social media analytics.

Kyung Min Park is an associate professor of strategy at the Yonsei University School of Business, Seoul, Korea.   He received his B.B.A. from Seoul National University, M.Sc. from Korea Advanced Institute of Science & Technology (KAIST), and Ph.D. from INSEAD, Fontaineblue, France.   His research interests include dynamics of strategic change and learning, competition in Internet market, and application of big data for strategic decision making. He has been publishing papers on strategic issues at international journals such as Asia Pacific Journal of Management., Organization Science, and Technological Analysis & Strategic Management.

Hae Chang Rim is a professor of computer science at the Korea University School of Computer, information and communication, Seoul, Korea.   He received his B.B.A. from Korea University, M.Sc. Missouri State University and Ph.D. from Texas State University, USA.   His research interests include Web Search, Network and Natural Language Programming. He has been publishing papers on engineering issues at international journals such as Personal and Ubiquitous Computing, Journal of Intelligent Information systems.