Extracting Urban Lifestyles Through Detailed Large-scale Purchase Records

Minjin Lee[1], Hokyun Kim[2] Bogang Jun[123], and Jaehyuk Park[4]

[1] Research Center for Small Business Ecosystem, Inha University, Incheon, Republic of Korea
[2] Department of Data Science, Inha University, Incheon, Republic of Korea
[3] Department of Economics, Inha University, Incheon, Republic of Korea
bogang.jun@inha.ac.kr
[4] KDI School of Public Policy and Management, Sejong, Republic of Korea
jpark@kdis.ac.kr

1 Introduction

Lifestyle encompasses an individual's way of life, including various aspects such as social status, consumption habits, and cultural interests. It is a multifaceted concept involving complex interactions among these dimensions, making it challenging to measure and capture comprehensively [1]. Recently, there has been a growing consensus that a more direct approach to understanding people's lifestyles is through the observation of their consumption patterns, which offer explicit insights into their choices, such as what they eat, how they spend their time, and the interests they prioritize [1–3].

The availability of large-scale purchasing records has enabled data-driven studies on consumption behavior, linked to consumer characteristics [3]. While credit card transaction records provide advantages in inferring purchase locations, their lack of information about detailed categories of consumed goods and services limits the ability to comprehensively capture latent lifestyle, which can be represented as a complex composition of purchased items related to socioeconomic status as well as individual preference [1, 4]. Differentiating spending on, for instance, golf gloves versus boxing gloves — which may signal distinct lifestyles and even socioeconomic statuses — remains challenging.

Here, we extract urban lifestyles using millions of detailed package delivery records from Seoul, Republic of Korea, which span four years. As a global metropolitan city with a diverse mix of social, cultural, and economic backgrounds, Seoul also stands as one of the world's largest e-commerce markets, with around 80% of its population using online shopping. Our dataset provides granular information on both recipient locations and the specific types of products purchased, allowing us to analyze urban lifestyles in unprecedented detail. Furthermore, we examine lifestyle variations across different urban areas, income brackets, and age demographics to explore the relationship between socioeconomic attributes and lifestyle patterns.

2 Data and Methods

Our main dataset consists of package delivery records spanning four years from 2018 to 2022, with a high spatial resolution (50 x 50 grid cells) and detailed categorization of

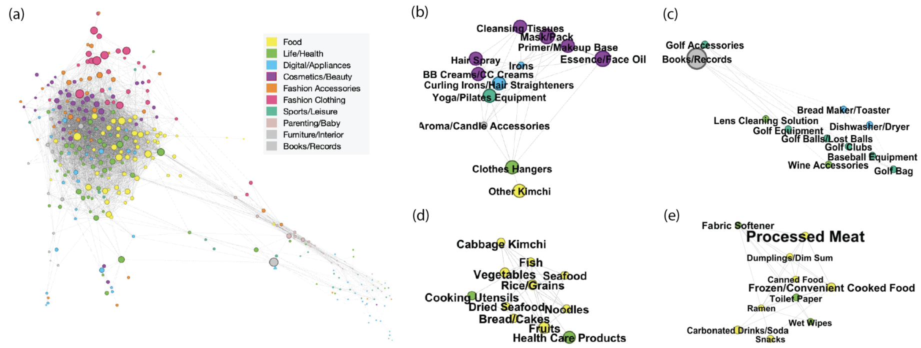

Fig. 1. Consumption space. (a) Product network based on co-purchase patterns. Node size represents the total consumption volume, and edge weight is calculated based on the proximity measure [5]. Colors represent the higher-level categories where each item belongs. (b)–(e) Ego networks for four example products: (b) Yoga/Pilates equipment, (c) Golf equipment, (d) Fruits, (e) Ramen.

the delivered items. By focusing on micro-level product categories, this study provides a more granular understanding of consumer behavior and lifestyle patterns.

We construct a product network based on being co-purchased by same grid cells to identify meaningful product baskets from the perspective of consumers. The weight between two products is calculated using the conditional probability that both products are consumed within the same grid cell, following the methodology of [5]. Since the original network is fully connected with noise, we extract the backbone of the network using the disparity filtering technique by [6].

3 Results

Fig.1(a) presents the network of products and their relationships, revealing the groups of products that are likely to be purchased together. The first notable finding is that products do not cluster according to their higher-level categories (e.g., food, cosmetics); instead, there is a significant overlap across different categories. This highlights a crucial insight: consumers tend to perceive and classify products based on their lifestyle and preference, rather than by predefined categories linked to basic functions. Moreover, this suggests that detailed information on co-purchased products is essential to extract urban lifestyle patterns through consumption behaviors.

Fig.1(b) to (e) show examples of how consumer baskets centered around two sports (Pilates and golf) and two food items (fruits and ramen) illustrate distinct lifestyles. Pilates-related purchases(Fig.1(b))are linked to cosmetics and self-care products, while golf-related purchases(Fig.1(c)) are associated with a broader range of hobbies like reading, music, and baseball. Fig.1(d) and 1(e) show dietary preferences. Fruit buyers often purchase healthy ingredients for home cooking, while instant ramen buyers prefer quick and convenient meals. Ramen consumption also correlates with household essentials like toilet paper and wet wipes, indicating a simple lifestyle typical of single-person households.

We further identified community structures within the network, revealing distinctive lifestyles as product baskets(Fig.2). The network is clustered into five primary communities at the optimal modularity level. The *Fashion-lovers* community is composed mainly of fashion clothing and accessories, while the *Beauty Enthusiasts* community encompasses various cosmetic products tailored for beauty purposes. The *Homemakers* community centers around food ingredients and household necessities, whereas the *Office Workers and Tech Enthusiasts* community features IT products, office supplies, convenience foods, and a range of drinks and health supplements. The *Parents and Hobbyists community*, distinct from the core of the network, includes various baby products and sports/leisure items, indicating a unique cluster focused on family-oriented and recreational activities.

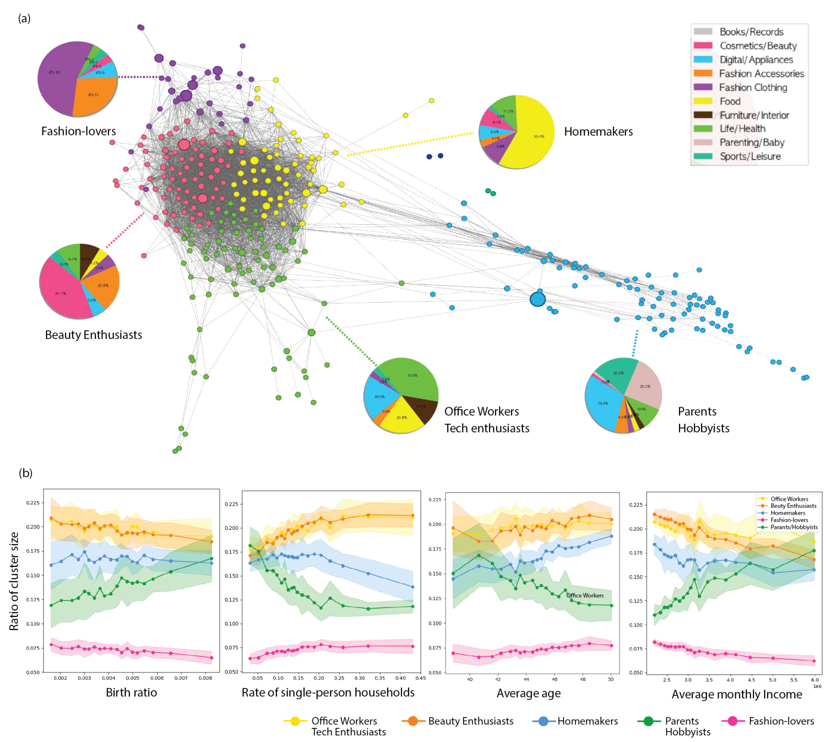

Fig. 2. Lifestyle communities and their relationship with socioeconomic attributes. (a) Community structure of the product network. Node color represents the community each product belongs to. Pie charts illustrate the composition of product categories. (b) Relationship between communities and socioeconomic indices — birth rate, rate of single-person households, average age, and average monthly income. The Y-axis represents the proportion of each cluster size at the district level, while the X-axis represents the corresponding socioeconomic attributes for each district.

We then analyzed the community composition at the district level, examining its relationship with socioeconomic characteristics (Fig.2(b)). While each category shows some correlation with certain socioeconomic attributes such as income level or age, no single socioeconomic index fully accounts for any specific lifestyle. For example, the Office Workers/Tech Enthusiasts, Beauty Enthusiasts, and Fashion-lovers communities are positively correlated with single-person households rate and negatively correlated with regional income. On the other hand, the Homemakers community is more prominent in regions having an older population and fewer single-person households. The Parents and Hobbyists community is also negatively correlated with single-person households but is more common in districts with higher incomes and birth rates. Our findings support the ongoing discourse in lifestyle studies, highlighting that lifestyle cannot be explained by a single social characteristic but instead reflects a complex interplay of various factors.

Summary. This study presents a novel approach to extracting urban lifestyles utilizing large-scale and detailed package delivery data, allowing for a more detailed examination of consumer behavior compared to traditional methods. By identifying co-purchase patterns and revealing consumer baskets, we uncover distinct lifestyle patterns across different socioeconomic groups and demographics, demonstrating how consumption patterns can serve as a significant proxy for understanding lifestyle choices. Our findings also indicate that the lifestyle groups do not align with any single socioeconomic characteristic. This provides strong evidence for the ongoing discussion about the complex relationship between lifestyle and socioeconomic factors, while also highlighting the effectiveness of our method as a better proxy for capturing the nuanced needs and preferences of diverse urban communities. The insights from our findings offer valuable guidance for policymakers. We expect that understanding how urban residents navigate their consumption choices based on their lifestyles can help policymakers develop and tailor public services, promoting inclusiveness and enhancing the quality of urban life.

References

1. DellaPosta, Daniel., Shi, Yongren. and Macy, Michael. : Why do liberals drink lattes? American Journal of Sociology 120.5, 1473-1511 (2015)
2. Zukin, Sharon.: Urban lifestyles: Diversity and standardisation in spaces of consumption. Urban studies, 35(5-6), 825–839 (1998)
3. Di Clemente, Riccardo., Luengo-Oroz, Miguel., Travizano, Matias., Xu, Sharon., Vaitla, Bapu., González, Marta C.: Sequences of purchases in credit card data reveal lifestyles in urban populations. Nature communications, 9(1), 3330 (2018)
4. Kozlowski, Austin C., Matt Taddy, and James A. Evans.: The geometry of culture: Analyzing the meanings of class through word embeddings. American Sociological Review, 84.5, 905-949 (2019).
5. Hidalgo, César A., Klinger, Bailey., Barabási, A.-L., Hausmann, Ricardo.: The product space conditions the development of nations. Science, 317(5837), 482–487 (2007)
6. Serrano, M. Ángeles, Marián Boguná, and Alessandro Vespignani.: Extracting the multiscale backbone of complex weighted networks. Proceedings of the National Academy of Sciences, 106.16, 6483-6488 (2009)