

THE ROLE OF CHANNEL CHARACTERISTICS ON BRAND-RELATED
USER GENERATED CONTENT

BY

MINGTAO GAO

THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of
Arts in Computational Social Science in the University of Chicago, 2021

Chicago, Illinois

Advisor:

Professor Jon Clindaniel

ABSTRACT

Using sentiment analysis, k-means clustering, and latent Dirichlet allocation (LDA), we collected and analyzed more than 50,000 user-generated content (UGC) related to five athletic apparel brands, to examine the content characteristics in brand-related UGC from three types of social media: Twitter (self-media), Reddit (collaboration platform), and Instagram and YouTube (creative outlets). We find that UGC on creative outlets contains more positive brand sentiments and emotional trigger words. UGC on self-media and collaboration platforms are brand-centered, while the most popular topics on self-media are brand news/trends, and it is a product-related question on collaboration platforms. UGC on creative outlets are self-promoting, while creators' motivations and achievements are the most popular on Instagram, and activity-related content and informative reviews are the most popular on YouTube. Besides, the formation of brand communities positively influences consumer engagement on social media. This research extended from the previous framework for comparing brand-related UGC with big data and machine learning techniques. It provides insights into how channel characteristics result in differences in brand-related UGC, which are valuable to future researchers and brand managers.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
CHAPTER 2: LITERATURE REVIEW	4
2.1 Social Media - Definition, Categorization, Psychological Motivations	4
2.2 Brand-Related UGC via Social Media	8
2.3 Research on Content Characteristics of UGC	11
2.4 Research on Channel Characteristics of UGC	13
2.4 Hypotheses	15
CHAPTER 3: METHODOLOGY	18
3.1 Data Collection	18
3.2 Data Analysis	21
CHAPTER 4: RESULTS & FINDINGS	23
4.1 Brand Sentiments	23
4.2 Textual Similarity	30
4.3 Topics Clusters	33
4.4 Topic Popularity	42
CHAPTER 5: DISCUSSION	53
5.1 Benefits	53
5.2 Limitations	57
5.3 Future Research	59
REFERENCES	61

CHAPTER 1

INTRODUCTION

Social media has become increasingly important in people's daily lives. As of January 2021, there were 4.66 billion active internet users worldwide, 59.5% of the global population (Statistic, 2021). Therefore, brand presence on social media becomes increasingly important to gain and improve brand recognition and communicate with brand audiences. In fact, the popularity of various types of social media has enriched the forms of communication and brand-consumer relationships (Gomez et al., 2019), and thus the sources of messages are no longer simply from the brands. Besides traditional paid media (paid or sponsored display advertisements), a brand's presence on social media allows them to publish brand-related content (owned media) and interact with users to reinforce their brand identities and give immediate responses to opposing voices. Moreover, social media allows consumers to control what brand-related information is generated, shared, and responded to; thus, consumers are connected and empowered like never before. They are no longer the receivers or endorsers of brand messages but the creators of brand information and brand value (Gomez et al., 2019) and the source of brand-related content, known as user-generated content (UGC) (Voorveld, 2019).

Brand-related UGC is vital to marketers, consumers, and researchers. First of all, previous empirical research papers have shown that UGC has significant effects on brand images, purchase intentions, and sales (Liu et al., 2017). UGC is often more influential than brand-created content in consumers' decisions because it reveals how real people use products and their experiences with the brand, which holds the authenticity of information about the brand or product. An online survey shows that 86% of consumers think "authenticity is important when deciding what brands they like and support", but 57% of consumers think "less than half of

brands create content that resonates as authentic" (Stackla, 2019). UGC, a new format of word of mouth, plays an essential role in consumers' purchasing decisions and consumers' views about the brands. In addition to the influences of UGC in consumers decisions, another reason UGC is vital to marketers is cost-efficiency. Because UGC is earned media, consumers voluntarily create assets for brands. For example, Google collected the best and funniest Nest Cam clips submitted by customers and made a campaign, which received much attention and popularity on social media with little cost spent. Lastly, because UGC can be tracked and measured to understand what motivates users to post brand-related UGC and what topics they like to discuss in their contents, they are valuable to social media researchers.

Although UGC is often considered as a single type of content, across different platforms and channels, various UGC on different social media platforms cannot be categorized as one channel for brand-consumer communications as a whole. In communication, channels refer to the medium that brings the message to its audience. Because the designs of social media platforms create diverse user experiences and interactions, many conceptual and theoretical research papers have tried to categorize and identify different types of social media and their characteristics. Kietzmann et al. (2011) present a honeycomb framework to define social media with seven functional building blocks. Each block represents a specific facet of social media functionality and corresponding user experience, and each social media platform achieves some or all of these building blocks to some extent. On the other hand, Zhu and Chen (2015) categorized four types of platforms based on the nature of connection (profile-based/content-based) and level of customization of messages (broadcast/customized): relationship media (e.g. Facebook, WhatsApp), self-media (e.g. Twitter, Weibo), creative outlets (e.g. YouTube, Instagram, Flickr), and collaboration platforms (e.g. Quora, Reddit). Two types of

categorizations (functions and user experiences vs. connectivity and types of messages) complement each other in comparing and distinguishing between various forms of social media. Based on the assumption that social media characteristics contribute to different user engagement experiences, Voorveld et al. (2018) conducted a survey study to measure user engagement with various social media across 11 experience dimensions and showed that “digital engagement is highly dependent on the platform.” For example, Twitter scored on topicality and negative emotion related to content, indicating Twitter ensures users are quickly informed and up to date with trends and news, but many contents induce negative emotions, while YouTube scored highest on entertainment and lowest on negative emotions (Voorveld et al., 2018). The differences in digital engagements can lead to distinct patterns of brand-related UGC on various types of social media, which has not been fully explored by prior literature.

This study aims to explore how the role of channel characteristics (connectivity, customization of messages, functionality, and user needs satisfaction) affects consumer creations of brand-related messages and consumer responses to brand communication in social media. *How do brand-related user-generated contents (UGC) differ across theoretically categorized social media platforms?* Specifically, we intend to examine whether and how brand-related UGC and user engagements differ across four social media platforms: Twitter (self-media), Instagram (photo-sharing creative outlet), YouTube (video-sharing creative outlet), and Reddit (collaboration platform). We expect them to differ by topics and sentiments of content such that brand sentiments on creative outlets are more positive than other channels, content on self-media and collaboration platforms are highly brand-focused, and users interact the most with content related to opinions on self-media, related to brand products on collaboration platforms, and related to experience on creative outlets.

CHAPTER 2

LITERATURE REVIEW

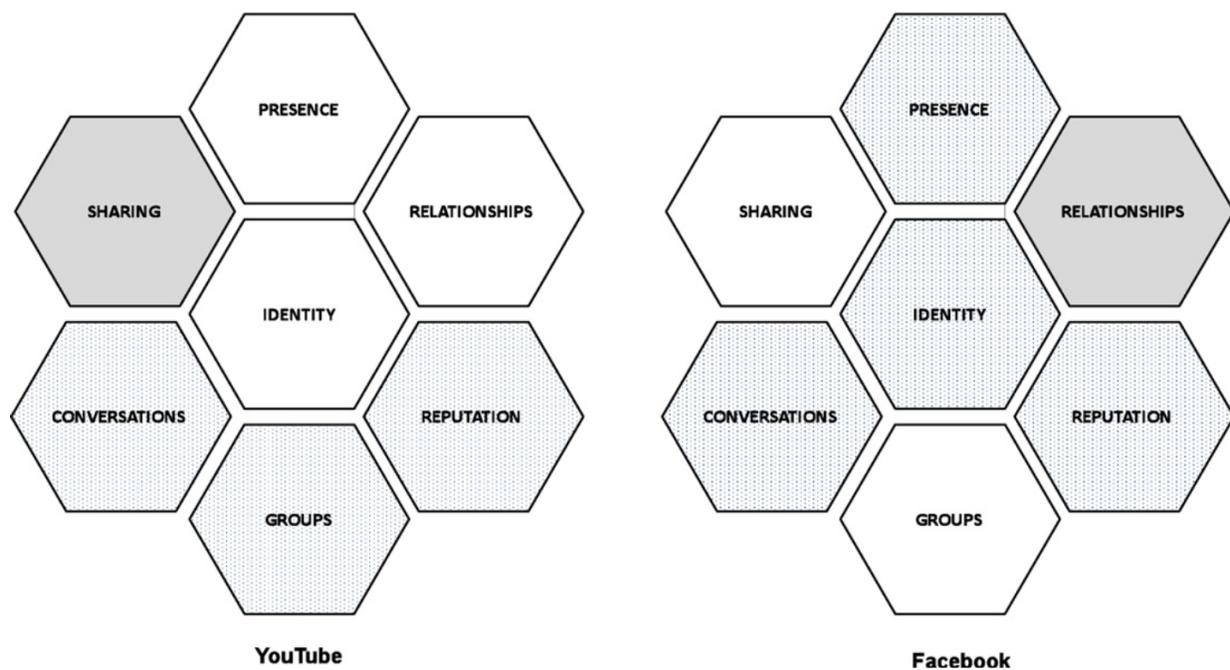
1. Social Media - Definition, Categorization, Psychological Motivations

Social software, the technical name for social media, was first defined by Shirley as “software that supports group interaction”. According to Dron (2007), the distinction between social software and earlier forms of mediated communications, like mailing lists and discussion forums, is its ability to reach a massive number of users. By the beginning of 2021, Facebook, YouTube, and WhatsApp are the top three most used social media applications, with more than 2000 million users each (Statistica, 2021). Other applications, including Instagram, TikTok, Reddit, and Twitter, also invite millions of users to communicate, create, interact, and share information. To understand the nature and structure of social software, Smith (2007) proposed the “social software honeycomb”, which illustrates seven building blocks of any social system: identity (how much users reveal their identities in a social media setting), communications (how much users can communicate with each other), sharing (the extent to which users exchange, distribute, and receive content), presence (the extent to which users can know if other users are accessible or online), relationships (how users can be related to other users, for example, Facebook friendship), reputation (whether users can identify the standing of others, including themselves), and groups (the possibility to form communities and sub-communities) (Smith, 2007; Pereira et al., 2010; Kietzmann et al., 2011).

Although almost all social media applications are built upon the seven functional blocks, the design and implementation of social software emphasizes different functionalities or achieves the same function in various ways. For example, to demonstrate the content’s reputation, YouTube uses the view counts or the number of likes received to show the reputation of a video,

while on Facebook, the number of likes is used. Besides, Facebook emphasizes the relationship between users and YouTube focuses on the contents created by users. One of the characteristics developed by Zhu & Chen (2015) to distinguish social media is the nature of connections, in other words, whether the application is profile-based or content-based. Profile-based vs content-based connection divides social media into two distinct groups with different purposes and natures of information (Zhu & Chen, 2015). Such distinctions can also be found in the honeycomb frameworks for YouTube and Facebook (Figure 1, Kietzmann et al., 2011). The primary functionality of YouTube is sharing, which leads to its emphasis on conversations and groups. Nevertheless, the primary function of Facebook is relationships, and thus Facebook emphasizes users' identity, presence, and reputation.

Figure 1. Honeycomb frameworks of YouTube and Facebook, the darker the color of a block, the greater this social media functionality is within the site (Kietzmann et al., 2011)



Moreover, Zhu and Chen (2015) categorized social media platforms into four types based on the nature of connection (profile-based vs content-based) and level of customization of

messages. The level of customization in the context of social media indicates the level of intended audiences of a message posted by users, whether the messages posted are customized for a specific person or group or broadcast to everyone on the platform. Therefore, four types of social media are relationship media (profile-based and customized message), self-media (profile-based and broadcast message), collaboration platforms (content-based and customized message), and creative outlets (content-based and broadcast). Examples of relationship media are Facebook, LinkedIn, WhatsApp. Self-media examples are Twitter and Weibo (Chinese version of Twitter). Collaboration platforms are Quora and Reddit. Lastly, creative outlets include YouTube, Instagram, Flickr, and Pinterest. Other than these categorizations of social media, these platforms also differ on other characteristics, for example, modality, privacy parameters, type of connections, and longevity of content accessibility, which all contribute to different engagement experiences among social media platforms (Voorveld et al., 2018).

Table 1. Types of social media and human need satisfaction (Zhu and Chen, 2015)

	Relationship	Self-media	Creative outlet	Collaboration
Poster	Relatedness, Self-esteem	Popularity-influence, Money-luxury*	Self-actualization, Competence	Competence, Self-actualization
Follower	Relatedness	Self-esteem, Relatedness	Pleasure-stimulation	Autonomy

In addition to studying social software functionalities and categories, researchers are interested in the psychological needs satisfied by those functionalities of different social applications. Based on prior literature, Sheldon et al. (2001) summarized the ten most fundamental needs satisfying human nature: autonomy, competence, relatedness, self-esteem, physical thriving, security, self-actualization, popularity or influence, pleasure or stimulation, and money or luxury. Autonomy, competence, and relatedness are three innate psychological needs, postulated by Ryan & Deci (2000) from the self-determination theory. Those needs are

salient in people's self-motivation and mental health, and they have event-related affect, indicating they are satisfied in daily activities (Ryan & Deci, 2000; Sheldon et al., 2001). In this case, the core human need that is satisfied by social media is relatedness. Activities that contribute to a general sense of relatedness and can be achieved through social media include communicating about personally relevant matters, participating in shared activities, feeling understood and appreciated, and participating in pleasant or otherwise enjoyable activities (Reis et al., 2000; Zhu and Chen, 2015). Based on an extensive review of prior literature, Zhu and Chen (2015) listed different human needs that are satisfied by four types of social media services they proposed (relationship, self-media, collaboration platform, creative outlets) (Table 1). Relationship media (e.g. Facebook) emphasizes creating relationships with other users and building one's identity, satisfying users' relatedness and self-esteem. Self-media (e.g. Twitter) satisfies users' needs for popularity/influence and money/luxury because the functionality of self-media focuses on building one's reputation. On the other hand, because the discussions and interactions focus on contents on content-based media, creative outlets (e.g. YouTube, Instagram) and collaboration platforms (e.g. Reddit) satisfies creators' needs for self-actualization and competence. Whereas, the difference between creative outlets and collaboration platforms is the need satisfied by the followers or viewers of content. On creative outlets, viewers look for aesthetic and entertaining content, satisfying their needs for pleasure. On collaboration platforms, viewers look for answers or help or try and learn new things, which is driven by the need for autonomy.

The prior literature shows social media platforms can be categorized into distinct types based on their architecture, functionalities, connectivity, and customization of messages, bringing users with different engagement experiences, and satisfying their different psychological needs.

At the same time, users visit these platforms with different intentions, interact with others in diverse ways, and generate content that looks unique from one site to another (Smith, Fischer & Yongjian, 2012).

2. Brand-Related UGC via Social Media

Various types of social media platforms facilitate new dynamic networks of consumers and brands and a wide variety of ways to easily create and share brand-related content, which gives consumers influence and credibility and has strong potential to shape brand perceptions and purchase behaviour (Alhabash, Mundel, and Hussain, 2017, p259). Creative consumers, those who produce value-added content in social media, are the new locus of value in the marketing of the new era (Berthon et al., 2012; Gensler et al., 2013). Sharing the experiences with, or opinions on, products and services with a multitude of other consumers allows those consumers to engage in electronic word-of-mouth (eWOM) communication (Hennig-Thurau et al. 2004). When UGC is brand-related, it overlaps with eWOM in its messages and influences on the brand.

Brand-related UGC is defined as content that “is made available through publicly accessible transmission media such as the Internet, reflects some degree of creative effort, and is created for free outside professional routines and practices” (Christodoulides, Jevons, & Bonhomme, 2012).

Consumers engagement on social media can be understood through the Uses and Gratifications (U&G) framework and COBRA (consumers’ online brand related activities). U&G, a widely used framework in media research, answers why and for what people use media (Khan, 2017). Shao (2009) and Muntinga, Moorman, and Smit (2011) introduced COBRA (consumer’s online brand related activities) to understand consumers’ behavioral engagement with brands on social media in three dimensions of analysis: consumption, contribution, and creation. Consumption of brand-related contents fulfills consumers’ information, entertainment,

and mood management needs; contribution through interacting with the content and other users to enhance consumers' social connections and virtual communities; creation of brand-related content achieves the need for self-expression and self-actualization (Shao, 2009). In this study, we focus on consumers' contribution and creation of brand-related contents on social media.

Hennig-Thurau et al. (2004) categorized reasons behind sharing brand-related UGC or engaging in eWOM communication into the ten fundamental needs that are satisfied by social media services: consumers' desire for social interaction (relatedness), desire for economic incentives (money-luxury), and the potential to enhance their self-worth (self-actualization). Similarly, Muntinga (2013) conducted interviews to study consumer motivations for engaging with brands and brand name products on social media, which are desires for information (autonomy), entertainment (pleasure), personal identity (self-esteem), social interaction (relatedness), and remuneration (money-luxury). Their research demonstrates that intrinsic motivations play an essential role in generating brand-related UGC behaviour. Intrinsic motivations (e.g. self-esteem, self-actualization, relatedness, competence) are driven by strong interests and involvement in the activity itself. In comparison, extrinsic motivations (e.g. money-luxury) are driven by the prospect of gaining some kind of extrinsic reward (Ryan & Deci, 2000). Additionally, the ego-defensive and social functions of attitude were found to have the most explanatory power in the user and creation of UGC, which implies that consumers can minimize their self-doubts and feel a sense of community in creating UGC (Daugherty & Bright, 2008). Engaging in such actions to connect with others and feel important reflects the intrinsic motivations for creating UGC.

Besides research on the underlying motives for brand-related UGC, many studies have explored the influence of brand-related UGC on brand image and potential sales. Varying

degrees of message valence and extremity (positive or negative) was found to influence consumer attitude toward the brand, and, as a result, positive brand-related reviews provokes consumer eWOM behaviour, brand engagement, and potential brand sales (Lee, Rodgers, & Kim, 2009). Besides, Kim and Johnson (2015) examined the influence of positive brand-related UGC shared via Facebook by asking participants to browse mock Facebook fan pages, and they found brand-related UGC activated the consumer's emotional and cognitive responses, which positively influenced behavioural outcomes in relation to the brand (i.e. brand engagement, potential brand sales). An online survey also indicates that consumer perceptions of co-creation, community, and self-concept positively impact brand-related UGC involvement, affecting consumer-based brand equity (Christodoulides et al., 2012). Dhar and Chang (2009) incorporated UGC, from blogs and social networking sites, in predicting future sales in the music industry and found a positive correlation between sales of an album and the volume of blog posts about the album. The creation of UGC can affect consumers' purchasing decisions by engaging them in actively elaborating a personal goal that is brand-related (Malthouse et al., 2016). Besides, the comparison between content from consumers (UGC) and marketers (marketer-generated content, e.g., MGC) reveals UGC exhibits a more substantial impact on consumer purchase behaviours through embedded information and persuasion than MGC (Goh, Heng, & Lin, 2013). Specifically, UGC positively influenced both brand equity and brand attitude, whereas MGC affected only brand attitude (Schivinski & Dabrowski, 2016). Therefore, brand-related UGC plays an important role in shaping consumers' brand perceptions, building and sustaining brand equity, and increasing future sales.

Since social media platforms have empowered consumers to control what information is generated and shared, there is significant uncertainty of what messages are produced by

consumers on which platform. McGuire (1989) proposed the Communication-Persuasion Matrix, which characterizes a set of components that can shape consumers' responses toward brand communication: source, message, channel, receiver, and destination. In this study, the source of communication is the consumers or the users on social media. We are interested in how posting on different types of social media platforms (channel characteristics) can influence the variations of features in the messages (message characteristics).

3. Research on Content Characteristics of UGC

Many researchers have examined the characteristics of brand-related UGC on social media. From marketers' perspective, UGC varies across a spectrum of creativity: from informal discussions about products and services, to structured reviews and evaluations in text or video, to self-created advertising videos (Pitt, Berthon, & Campbell, 2008), to finally, the modification of proprietary products and services and the distribution of those innovations (Berthon et al., 2012). Based on prior literature, O'Hern & Kahle (2013) defined four types of UGC along two key dimensions: users' contribution objective (product promotion vs product innovation) and direction of UGC knowledge flows (C2C knowledge row vs C2B knowledge row). Informing UGC (product promotion and C2C knowledge row) is the most pervasive content we find on social media, through which consumers target their fellow users by providing them with ideas, insights, and thoughts on a product or brand and express their satisfaction (or dissatisfaction).

One form of informing UGC that has been extensively studied is review. Positive reviews can boost potential sales, for example, the number of bookings on a travel website (Ye, Law, Gu, & Chen, 2011), and credibility of the source plays an important role in consumers' attitude toward using UGC (Ayeh, Au, & Law, 2013). Moreover, the pervasiveness of sponsored content also yields comprehensive related studies. Sponsored content is defined as "the intentional

incorporation of brands, products, or persuasive messages into traditionally non-commercial, editorial content” (Boerman, Reijmersdal, Neijens, 2014). Organic (unpaid or not-sponsored) and experience-centric content is found to be more favourable than sponsored UGC (Kim and Song, 2018). Van Reijmersdal et al. (2016) found that sponsored content in blogs induces both cognitive (counterarguing) and affective (negative affect) resistance strategies to decrease the persuasion attempt made in the blogs. Thus, the current research implies that sponsored content has a more negative influence on brand attitude than other forms of brand-related UGC (Müller & Christandl, 2019). Sponsored content on Twitter has also raised ethical concerns because consumers can hardly distinguish between sponsored and non-sponsored tweets, leading them to be vulnerable to persuasion in the sponsored contents (Kim & Song, 2018; Gomez et al., 2019).

However, prior literature about the UGC on social media is generated mainly from experiments and surveys (Kim and Song, 2018; Muntinga et al., 2011; Pentina et al., 2018; Müller & Christandl, 2019). Social media is unique with regard to the massive amounts of data they provide, and thus the use of actual social media data is necessary to examine and explain brand-related consumers’ activities in a natural setting (Voorveld, 2019). The big data of UGC on social media shows its advantage in applying machine learning methods in UGC-related research. Netzer et al. (2012) deployed a text-mining approach with semantic network analysis tools to convert UGC into market structures and competitive landscape insights. The combination of marketing methods and text mining allows automatic extraction of product attributes and relative brand positions from online consumer reviews (Lee and Bradlow, 2011). Further, the clustering of movie reviews can be used to predict box-office sales (Moon, Park, and Kim, 2014). Sentiment analysis, latent Dirichlet allocation (LDA), and text mining on UGC on social media are general approaches to extract product dimensions and measure brand sentiments

(Schweidel and Moe, 2014; Tirunillai and Tellis, 2014; Homburg, Ehm, and Artz, 2015; Ma, Sun, and Kekre, 2015). Moreover, Liu et al. (2017) applied text mining with LDA and sentiment analysis on brand-related UGC from Twitter for 20 brands across five industries. Product, service, and promotions are dominant topics consumers discuss the brands on Twitter and sentiments toward brands vary within and across industries (Liu et al., 2017). This paper provides a framework for deriving brand-related topics and sentiments to transform social media data into brand insights. However, as their focus is on topics and characteristics of brand-relative UGC, the paper fails to compare how UGC varies across multiple platforms systematically. The limitation of platforms studied in previous research papers is pervasive. Voorveld (2019) found that 30% of all social-media-related papers published between 2008 and 2018 focused primarily on Facebook, and she argued that future research should more frequently compare different social media platforms.

4. Research on Channel Characteristics of UGC

Each type of channel provides a unique medium to bring messages to its audience, and communication channels directly influence what kinds of discussion consumers have about brands and products (Thorson and Rodgers, 2012; Berger and Iyengar, 2013). Cross-platform studies on consumer-generated reviews have discovered distinct patterns of reviews. Specifically, consumers left more reviews and longer reviews at Amazon.com, compared to Barnesandnoble.com (Chevalier and Mayzline, 2006), and left more positive or promotional reviews for hotels with a high incentive to fake on TripAdvisor than Expedia (Myzlin et al., 2014). These studies shed light on the topics for brand-related UGC, and consumer responses to brand communication may differ across different types of social media platforms (i.e. relationship media, self-media, creative outlets, collaboration platforms).

Since different functionalities and characteristics of social media platforms translate into different consumer experiences, a consumer's engagement with advertising is highly dependent on the platform they use (Voorveld et al., 2018). Smith et al. (2012) conducted a cross-platform study of brand-related UGC by sampling 600 UGC posts for two retail-apparel brands on YouTube, Facebook, and Twitter. They analyzed those posts across six dimensions (promotional self-representation, brand centrality, marketer-directed communication, response to online marketer action, factual information about the brand, brand sentiment), leading to exciting differences and consistencies between brand-related UGC three social platforms. Brand-related UGC on YouTube was more likely to feature consumer self-promotion and least likely to be brand-focused than on Facebook or Twitter (Smith et al., 2012). Expanding on the six dimensions, Roma & Aloini (2019) incorporated six new dimensions of UGC in their study: response to advertising campaigns, location sharing, connection with personal experience, real-time sharing of purchase experiences, real-time sharing of consumption experiences, and exhibition of brand recommendation features. Using the same methods as Smith et al. (2012), Roma & Aloini (2019) collected UGCs for each of the two brands, Coke and Pepsi, on three social media platforms (i.e. Facebook, Twitter, Instagram). Both studies provide a framework for comparing brand-related UGC across three particular social media channels, but data collected and brands analyzed in this study are limited. Thus, platform-based studies fail to extrapolate how different functionalities of each social media platform lead to distinct patterns exhibited in brand-related UGC. New social media platforms are constantly being created, yet the core functionalities of new platforms are built based on seven fundamental blocks. Thus, further research needs to consider the characteristics of each platform in influencing consumer responses and marketers' use of the platforms (Voorveld, 2019).

Another challenge concerning channel characteristics is that social media is usually not used as a stand-alone activity in brand communication but is combined with other media (Voorveld, 2019). Voorveld et al. (2018) found that advertising offline media (TV, radio, newspapers, magazines, out of home) is relevant in driving consumers' engagement with brands on social media, yet consumers more often than not use social media in combination with other digital and traditional media in their consumer journey (Kannan and Li 2017; Voorveld, Araujo, et al. 2018).

In all, previous research on motives and content characteristics of brand-related UGC, using big data and text analysis, lacks cross-platform comparison (Liu et al. 2010; Jaakonmäki, Müller, and Vom Brocke, 2017). Whereas previous research on channel characteristics focused mostly on a single platform and drew conclusions from small surveyed samples without using the big data of UGC (Kim and Song, 2018; Muntinga et al., 2011; Pentina et al., 2018). Besides, cross-platform studies fall short of discussing how the functionalities and psychological needs satisfied by those platforms may influence different content characteristics (Smith et al., 2012). Therefore, this research aims to use the big data of brand-related UGC and theoretically categorized social media platforms to investigate how channel characteristics influence consumer-generated content and consumer responses to brand communication.

5. Hypotheses

a. H1- Brand sentiments in UGC on creative outlets are more positive.

Previous studies show UGC brand sentiments are different across social media sites, and a significant positive sentiment is observed on YouTube, compared with Twitter and Facebook (Smith et al., 2012; Roma & Aloini, 2019). Smith et al. reasons for such a distinction are due to YouTube's culture of self-promotion. However, from a psychological perspective, users on

YouTube, a creative outlet, post content for self-actualization and view content for pleasure. Thus, incorporating positive brand sentiments is one way to make content more entertaining and visually pleasing. Moreover, although Twitter and Instagram are both built on the functional blocking of sharing, sharing on Twitter leads to the formation of conversations via tweets and retweets, while sharing on Instagram is a way to build user reputations. When contents are created to build conversations, we expect a mix of negative and positive voices, which explains why the percentage of negative tweets is much higher than that of positive tweets among all brands on Twitter (Liu et al., 2017). Therefore, we hypothesize the UGC on creative outlets (i.e. YouTube and Instagram) demonstrate more positive sentiments than content on other platforms, and conversation-based self-media (i.e. Twitter) tends to generate more negative UGC.

b. H2 - UGCs on self-media and collaboration platforms are more brand-focused than those on creative outlets.

Brand centrality refers to how much brand is discussed in the brand-related UGC, in other words, whether the brand is the focus of the content or is featured peripherally. UGC is most likely to be centered around the brand on Twitter and least likely on YouTube (Smith et al., 2012). We expect the same pattern to be found in this study. Besides, since groups and collaborative conversation are two main functional blocks of Reddit, which addresses users' needs for information, we expect UGC on Reddit is also more likely to be central to brands. Common brand-centered topics in UGC on Twitter are related to product, service, promotion, competitors, news/trends, and price (Liu et al., 2017). Because product and promotion are found to be discussed most often (55.9% and 15.3%) on Twitter, regarding UGC related to brands in the footwear industry (Liu et al., 2017), we expect content on Twitter and Reddit to be more brand focused, additionally, product and promotion will remain the most prevalent on Twitter.

Moreover, self-promotion is defined as an "effort to express a specific image and identity to others" (Żywica and Danowski, 2008), and consumers often use brands as symbols to construct their online images (Smith et al., 2012). Brand-related UGC on YouTube is found to feature consumer self-promotion (Smith et al., 2012). Because YouTube and Instagram, as creative outlets, satisfy consumers' needs for self-actualization and competence by creating brand-related content, we hypothesize that content on creative outlets focuses more on the creator's personal experiences or daily life.

c. H3 - The most popular contents are opinion-based on self-media, product-focused on collaboration platforms, and experience-based on creative outlets.

Looking at different human needs satisfied by followers on each type of social media (Table 1), we found relatedness is the top psychological need satisfied by Twitter, users on Instagram and YouTube look for contents that provide them with pleasure or stimulation, and users on Reddit are attracted to content with helpful information (Zhu and Chen, 2015). Therefore, we hypothesize that popular contents are opinion-based on Twitter, product-focused on Reddit, and contents about personal experiences or daily life receive the most popularity on creative outlets.

Which brand-related UGC receives more popularity than others is dependent on brand images and the marketing strategies deployed. Compared to Puma and Under Armour, who build their brand images through sponsorship and endorsement and focus on promoting their products, Lululemon and GymShark actively try to create brand communities on social media platforms. Consumers engaged in social media brand communities develop positive attitudes towards the brand, such as trust and loyalty (Habibi et al., 2016; Coelho et al., 2018). We expect to find more positive sentiment in content related to Lululemon and GymShark. Besides, consumers' intrinsic motives of altruism and social identification motivate them to engage in brand communities

(Lee, Kim, & Kim, 2011). Content related to Lululemon and GymShark receive more popularity when they are about experiences and motivations that are helpful to others because of the sense of community formed by the two brands.

CHAPTER 3

METHODOLOGY

1. Data Collection

To answer the research question of whether and how brand-related UGC differs across types of social media, we need to decide which platforms to extract user data from. Smith et al. (2012) collected UGC from three social media platforms (i.e. Twitter, Facebook, and Youtube) because they are “the most popular sites in terms of consumer usage and marketer interest”, and three platforms differ in the design and functionalities: a microblog (Twitter), a social network (Facebook), and a content community (YouTube). However, in this study, we focus not only on differences between design and functionalities among various social media platforms but also on consumers’ psychological needs satisfied by posting brand-related content on those platforms. Those needs are crucial to understanding distinct patterns of UGC across platforms.

Four social media platforms are chosen for this study: Twitter, Instagram, Youtube, and Reddit. First of all, four social platforms are categorized into three types based on the nature of connection and the customization of messages (Zhu and Chen, 2015): Twitter represents self-media, Instagram and YouTube are creative outlets, and Reddit represents a collaboration platform. Each platform differs in its design and functionality and the way it satisfies user needs. Twitter satisfies user needs for popularity and money, while the remaining three platforms satisfy creators’ needs for self-actualization and competence. Moreover, Instagram and YouTube both

satisfy viewers' needs for pleasure, but contents are delivered in two different forms (image vs. video), while Reddit satisfies users' autonomy needs to look for information or help.

Relationship media (i.e. Facebook) was excluded in this study because it emphasizes creating relationships with other users and building one's identity (needs for relatedness and self-esteem). Thus, consumers are less inclined to discuss brand-related content, which is supported by findings from Smith et al. (2012), that, compared to Twitter and YouTube, UGC collected from Facebook shows no significant pattern in all six dimensions measured.

In terms of brands, Smith et al. (2012) collected UGC from two apparel brands, Lululemon and American Apparel, to study whether a more proactive social media marketing strategy impacts brand-related UGC. Because consumers' sentiments and topics of interest vary across industries, even on one platform (Liu et al., 2017), we decide to focus on different brands from one industry - athletic apparel. Following two criteria proposed by Liu et al. (2017), brands must be owned by an S&P 500 company and be closely related to consumers' daily lives, and five leading athletic clothing brands are chosen for this study: Nike, Lululemon, Puma, Under Armour, and GymShark. All five brands provide similar products and services, but each deploys different marketing and social media strategies and creates a distinctive brand image.

Nike, the world's leading sports footwear and apparel company, targets consumers from 15-45. Nike focuses primarily on innovation and emphasis on research and development of new products (Mahdi et al., 2015). In recent years, because of increasing investments in digital marketing, Nike has become the leader in the digital market, garnering the highest number of fans and followers across four platforms. Nonetheless, statistics show that Puma is relatively far behind in the race with Nike and Adidas (Mahdi et al., 2015). In the past 6 years, Puma focused away from lifestyle and fashion towards retro sports products and footwear, leading to doubled

annual sales and operation profit (SportsPro, 2020). Because of Puma's mission statement, "to be the fastest sports brand in the world," Puma has slowly and gradually focused on influencer marketing over the past few years and it has become known for sponsoring many well-known football players to promote its brand. As a result, Puma's social media seems to concentrate more on presenting their products and sponsorship-related contents. Similarly, Under Armour invests mainly in sporting partnerships and celebrity endorsement deals to promote its products and heighten brand consideration among target consumers. Under Armour also uses social media platforms to push out product-related information rather than communicate with its customers.

On the other hand, brands that try to create a sense of community among fans and followers are usually more active than brands that majorly focus on sponsorships and endorsements. Lululemon and GymShark are two brands that actively engage in creating brand communities. Lululemon is a fast-growing yoga-inspired clothing brand founded in 1998. Instead of traditional advertisements, Lululemon uses word of mouth to promote its products, and it tries to create close relationships with its customers. The representatives of Lululemon are customers' yoga instructors or gym trainers. Similarly, fans of GymShark believe, rather than a brand, GymShark represents a supportive community that they can feel a part of (Inoue et al., 2020). GymShark started the partnership between brand and fitness influencers ahead of all other brands, making it the fastest-growing company in the UK within four years after it was founded. Besides, unlike Nike, Puma, and Under Armour, who try to reach both men and women between the ages of 17-45, Lululemon targets a smaller niche of yoga lovers and GymShark targets younger audiences of 18-25-year-olds. Due to the product prices, the majority of GymShark consumers are student-athletes, and Lululemon consumers are females with high disposable incomes.

We collected more than 50,000 brand-related UGC posts from four social media platforms. UGC from Twitter, Reddit, and YouTube is collected by public APIs, and we designed customized web crawling algorithms in Python to collect Instagram posts. Because of differences in design and functionalities of the four platforms, different scraping strategies are deployed to ensure the contents collected from each platform are brand-related. Reddit posts are collected from subreddit forums (e.g. r/lululemon, r/nike). On Twitter, any posts that contain the brand name (e.g. Nike, Puma) are collected. Similarly, YouTube videos that contain the brand name in the title and have scripts or English transcripts are collected. Because Instagram does not allow users to find posts by keyword, we scraped Instagram posts that contain brand hashtags (e.g. #lululemon, #underarmour). To comply with the ethical standards and preserve users' privacy, only publicly available posts are collected. Because of the inherent differences in the form of contents and limitations of APIs, we cannot keep UGC collected from each platform to be the same size, but we tried to keep the data size consistent among the five brands on each platform. Specifically, we collected around 5000 tweets from Twitter, 2500 posts from Reddit and 2500 posts from Instagram, and 500 video scripts from YouTube for each brand.

2. Data Analysis

Smith et al. (2012) analyzed UGC data from six dimensions, and each dimension was manually coded. Such a method was achievable for 600 UGC posts, and it is not feasible for studies on big data of UGC posts. Thus, this study used machine learning techniques, including sentiment analysis, LDA, and topic modelling (i.e. clustering), to analyze 50,000 UGC collected.

To answer the research question, the differences of brand-related UGC among three types of social media platforms, we analyzed language patterns at both word-level and text-level. With raw text data collected, we explored word-level text characteristics by examining the

distribution of terms and phrases and distinguishing important words from posts. To examine the differences in word-use pattern, we calculated and visualized differences (divergences and distances) between word frequency distributions of brand-related UGC from four platforms. Next, we moved from word-level exploration to discover higher patterns and topics in all posts from four platforms. First, clustering is a good machine learning technique to partition text data based on patterns of covariation among available text features, and we deployed four different clustering approaches: flat clustering, k-means, a hierarchical approach, and Ward's method. Two-dimensional content clustering (i.e. LDA) is used to discover the hidden themes and topics in each collection of posts. Then, we further analyzed those topics and explored differences in topics discussed on four platforms.

Moreover, we are interested in what topics received the most and least popularity on each social media platform due to the different psychological needs of viewers. The popularity of content can be measured by how many users consume (e.g. watch a video, read comments, and view likes/dislikes without responding) and participate/contribute to a brand-related UGC (Shao, 2009). However, because our study primarily focuses on how other users participate in any given content and the measurement of content consumption (views) is not available in many social media platforms, the popularity received by each UGC is measured by the number of likes it receives. "Likes on social media represent a form of user vote or an expression of appreciation of content. The number of likes indicates the popularity of content. (Khan, 2017)" Likes are measured slightly differently across four platforms: the number of upvotes on Reddit and the number of likes on Instagram, Twitter, and YouTube. Those attributes (upvotes and likes) represent how other users interact with the content and thus indicate other users' inclination or agreement towards the content, which is valuable to the brand. Finally, we explored what themes

and topics are discussed by content that received the greatest attention and lowest attention on each social media platform.

CHAPTER 4

RESULTS & FINDINGS

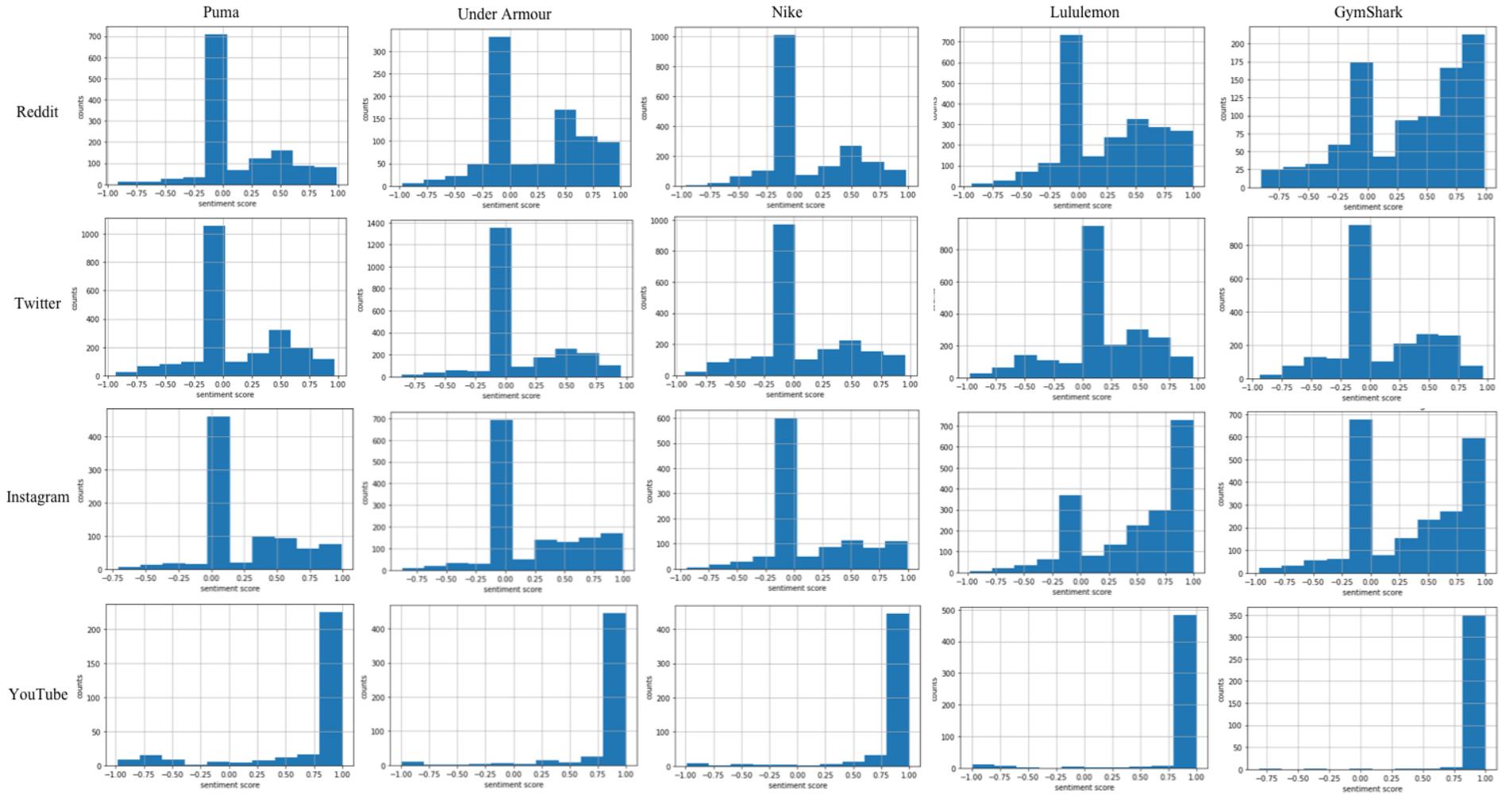
1. Brand Sentiments

Sentiment analysis, also known as opinion mining, computationally determines the affective state of a piece of text data as positive, negative, or neutral. It helps us to understand the opinion and attitude that users on various social media platforms have towards a specific brand. We calculated the sentiment score for each piece of UGC, computed the average sentiment score for each brand on four platforms (see Table 2), and plotted the distributions of their compound sentiments (see Figure 2). The compound score normalizes the sum of positive, negative, and neutral scores and thus ranges between -1 (most negative) and 1 (most positive).

Table 2. Average compound sentiment scores for 20 datasets (5 brands across 4 platforms)

	Reddit	Twitter	Instagram	YouTube
Puma	0.166	0.141	0.230	0.708
Nike	0.157	0.109	0.186	0.876
Under Armour	0.250	0.150	0.242	0.843
Lululemon	0.276	0.137	0.506	0.899
GymShark	0.350	0.123	0.381	0.968

Figure 2. Sentiment score distributions for 20 datasets (5 brands across 4 platforms)



The average sentiment scores and the distributions of sentiments support H1, that UGC brand sentiments are more positive on creative outlets (i.e. Instagram and Youtube). The average sentiment scores of Instagram and YouTube are higher than those of Reddit and Twitter across five brands. Contents on YouTube are observed to be significantly more positive than contents on the other three platforms. Although average sentiment scores reveal that brand-related UGC on Instagram contains more positive sentiments than those on Reddit, we do not find much of a difference between distributions of brand sentiments on Instagram vs. on Reddit.

Among all four social media platforms, brand sentiments on Twitter are the lowest, confirming our hypothesis that consumers on Twitter, a conversation-based platform, create brand contents that contain less positive sentiments. However, unlike the previous finding from Liu et al. (2017), that the percentage of negative tweets is higher than that of positive tweets, UGC on Twitter contains primarily neutral or slightly negative sentiments. Liu et al. (2017) concluded a brand sentiment is correlated with consumers' overall satisfaction with the brand, and they found "unhappy customers are about three times more likely to engage in negative eWOM than happy customers are to engage in positive eWOM" (Liu et al., 2017). Although many studies have shown dissatisfied customers are more likely to engage in WOM (Richins, 1983; Anderson, 1998), Twitter, as a micro-blog, is not considered a platform for unhappy customers to leave negative reviews (like Yelp). Our finding that most tweets contain neutral or slightly negative brand sentiments does not support overwhelmingly negative sentiments towards all brands found in Liu et al. (2017). Our finding also supports the core functionality of Twitter as a conversation-based channel rather than a review platform where dissatisfied consumers leave negative reviews. Moreover, Jansen et al. (2009) found, among brand-related tweets (19% of tweets), only 20% contained some expression of brand sentiments, which implies the brand is

not the primary focus of tweeting. Users on Twitter express their opinions or seek to provide information to form conversations through the creation of brand-related content, which also explains why brand-related UGC contain primarily neutral or slightly negative sentiments.

Surprisingly, brand sentiments on Reddit are comparatively more positive than we expected, which shows a similar pattern as Instagram. We found that consumers express more positive brand sentiments on content-based social media than profile-based sentiment. Both creative outlets and collaboration platforms can satisfy posters' needs for self-actualization and competence, as they provide users with an outlet to “showcase their creativity and receive recognition from fellow users” (Zhu and Chen, 2015). Self-actualization refers to a sense of long-term growth and the feeling of developing one's potential and making life meaningful, and competence refers to the feeling of capability in one's actions and performance (Shelden et al., 2001; Zhu and Chen, 2015). Those needs explain the motivations for content creators to discuss brand-related topics on content-based platforms, and thus, those contents tend to contain more positive feelings. Therefore, our first hypothesis is supported, and brand sentiments are observed to be more positive on content-based platforms (i.e. creative outlets and collaboration platforms).

Moreover, brand sentiments not only differ across platforms but also differ among brands. Contents related to Lululemon and GymShark contain significantly more positive sentiments than the other three brands. Although Liu et al. (2017) explain that differences in brand sentiments are caused by consumer satisfaction with brand products, we cannot simply conclude that Lululemon and GymShark provide better products and services than Nike and Under Armour. Lululemon and GymShark actively engage with consumers across social media platforms to build their brand communities, which leads to prominent numbers of positive brand sentiments in UGC related to the two brands. This result supports that consumers tend to develop

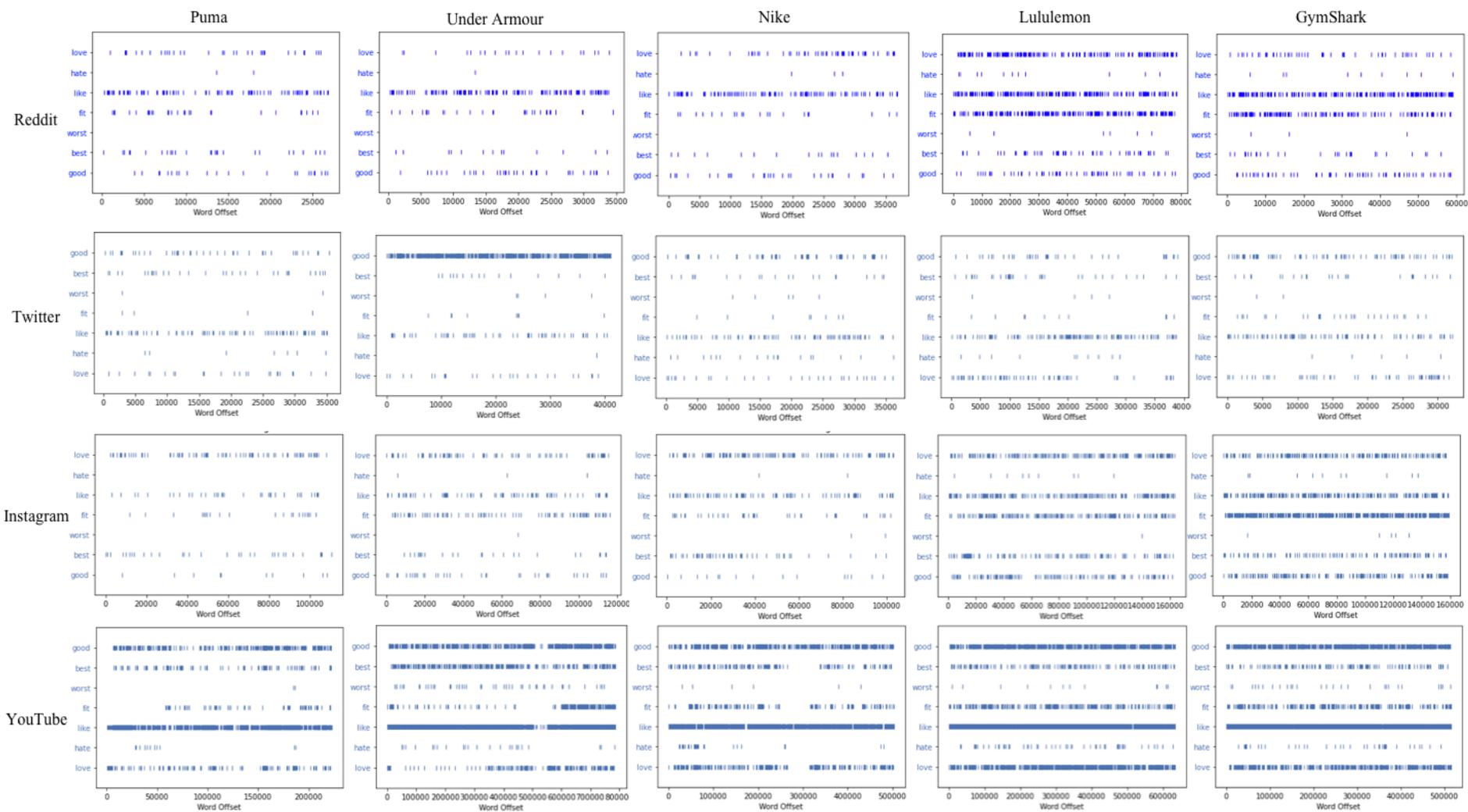
positive attitudes towards the brand by engaging in social media brand communities (Habibi et al., 2016; Coelho et al., 2018). Therefore, brand sentiments on social media are strongly influenced by whether the brand has created and owns an online brand community.

We have analyzed brand sentiments as either being positive or negative, yet emotions can be expressed in different degrees or levels of positive/negative feelings. For example, “love” is considered as a verb containing stronger positive feelings than “like”, and similarly, superlative adjectives (e.g. best and worst) are considered to contain stronger feelings than comparative adjectives (e.g. good and bad). To further explore how emotions are expressed across platforms, we calculated the frequencies and plotted the occurrences of a set of words that imply emotions and feelings at different levels (i.e. love, hate, like, worst, best, good) with respect to where in the text they appeared (see Table 3 and Figure 3).

Table 3. Frequencies of strong positive, strong negative, and weak positive words in 20 datasets.

Brand	Measurements	Reddit	Twitter	Instagram	YouTube
Puma	Strong Positive Words (best & love) Freq (%)	0.243	0.166	0.397	0.119
	Strong Negative Words (worst & hate) Freq (%)	0.008	0.290	0.000	0.008
	Weak Positive Words (good & like) Freq (%)	0.428	0.393	0.148	0.818
	All Emotion Words Freq (%)	0.679	0.588	0.545	0.945
Nike	Strong Positive Words (best & love) Freq (%)	0.213	0.150	0.469	0.118
	Strong Negative Words (worst & hate) Freq (%)	0.008	0.068	0.008	0.070
	Weak Positive Words (good & like) Freq (%)	0.409	0.502	0.235	0.872
	All Emotion Words Freq (%)	0.630	0.721	0.712	0.998
Under Armour	Strong Positive Words (best & love) Freq (%)	0.096	0.110	0.174	0.081
	Strong Negative Words (worst & hate) Freq (%)	0.003	0.012	0.007	0.010
	Weak Positive Words (good & like) Freq (%)	0.546	2.012	0.169	0.510
	All Emotion Words Freq (%)	0.645	2.142	0.350	0.602
Lululemon	Strong Positive Words (best & love) Freq (%)	0.316	0.273	0.262	0.235
	Strong Negative Words (worst & hate) Freq (%)	0.022	0.042	0.007	0.013
	Weak Positive Words (good & like) Freq (%)	0.422	0.441	0.278	1.750
	All Emotion Words Freq (%)	0.760	0.756	0.547	1.990
GymShark	Strong Positive Words (best & love) Freq (%)	0.177	0.296	0.243	0.149
	Strong Negative Words (worst & hate) Freq (%)	0.020	0.026	0.012	0.016
	Weak Positive Words (good & like) Freq (%)	0.561	0.617	0.308	1.683
	All Emotion Words Freq (%)	0.745	0.938	0.562	1.848

Figure 3. Word occurrences plots of 20 datasets (5 brands across 4 platforms). The list of words is “love, hate, like, fit, worst, best, good”. Each dot represents one occurrence of the word in the text dataset, and thus more dots indicate many occurrences of the word.



“Love” and “best” (strong positive emotion), “hate” and “worst” (strong negative emotion), and “like” and “good” (weak positive emotion) are emotional trigger words that consumers often use to describe their opinions on brands and products or discuss their personal experiences, but they differ in sentiments and levels of feelings contained.

The trends shown in the use of emotion words are consistent with our findings in brand sentiments. Content on platforms with more positive brand sentiments (i.e. YouTube) and related to brands with more positive sentiments (i.e. Lululemon and GymShark) make use of emotion words more frequently, as the frequencies of emotion words in all YouTube datasets are the highest compared to those of other platforms (around 2% of words are emotion words in Lululemon-related YouTube dataset). The highest frequencies indicate both positive and negative emotion trigger words appeared the most often on YouTube. Because UGC data scraped from YouTube are generated from video scripts, it is a form of recorded verbal communication. Different from text data on other platforms, verbal communication tends to be more informal and, to engage with the audience, YouTube creators often need to use more words to express themselves and trigger viewers’ emotions. Besides, if we compare the frequencies of words that indicate strong emotions (i.e. “love” and “best”) and weak emotions (i.e. “like” and “good”), UGC on Instagram are expressed more often in strong emotions while UGC on the other three platforms contains more weak emotion words. The plots also show that “love”, and “best” appear more frequently on YouTube and Instagram brand-related UGC than on Twitter. Besides, the examination of the most frequently used adjectives in each dataset shows “great” appears to be one of the top 10 adjectives used in all YouTube datasets, and “happy” appears to be one of the top 10 adjectives used in all Instagram datasets. This kind of pattern, one adjective frequently appearing in UGC related to all five brands, is not observed on other platforms (i.e. Twitter and

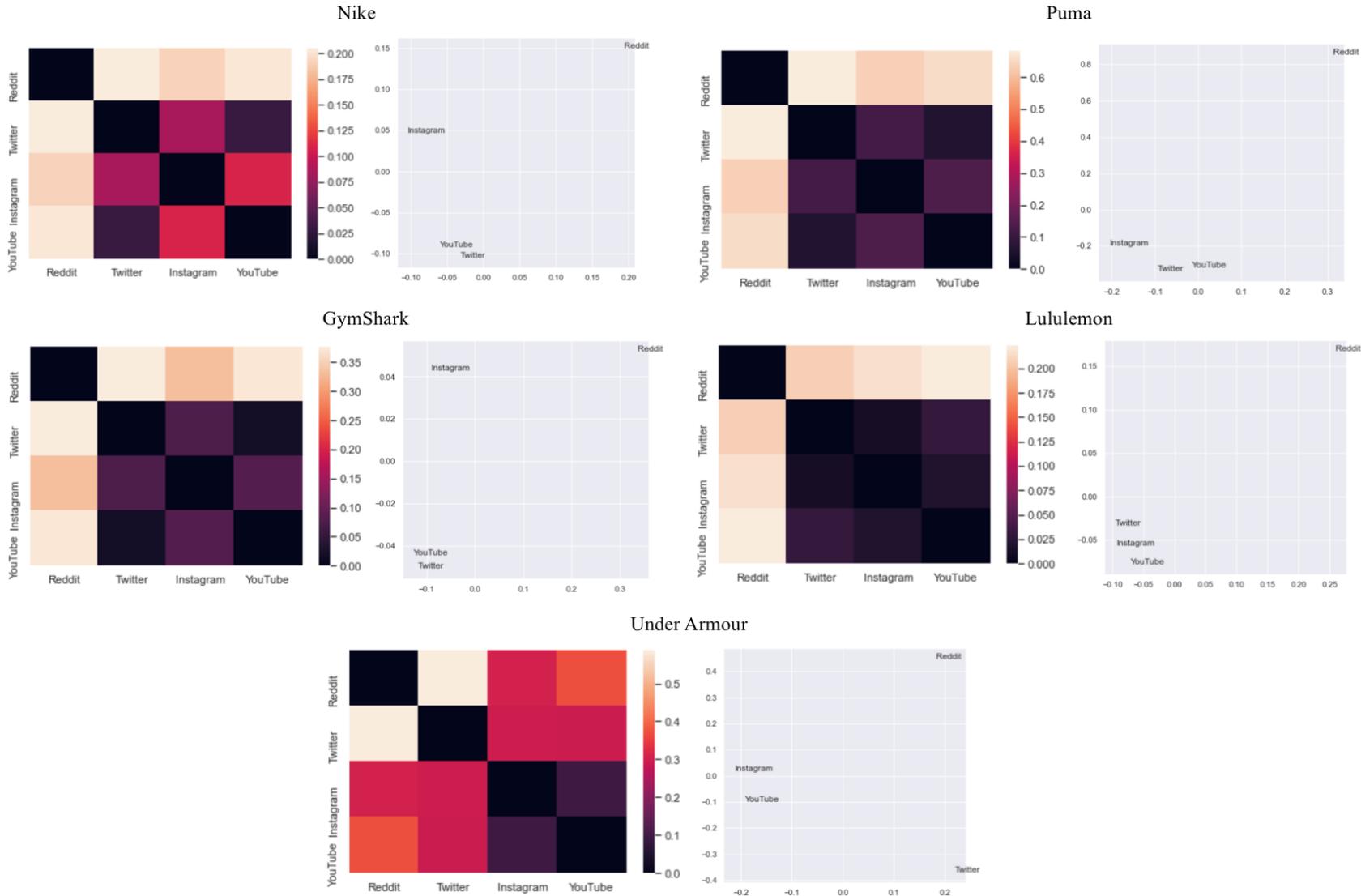
Reddit). Therefore, brand sentiments are found to be more positive on creative outlets, and consumers tend to use more emotional words when posting on YouTube and use stronger emotional words when posting on Instagram.

Additionally, if we compare frequencies of emotion words across five brands, Lululemon and GymShark-related contents contain more emotional trigger words than content related to other brands across four platforms. This is not surprising, since we have found that Lululemon and GymShark contain significantly more positive sentiments. Further, if we compare the use of strong vs weak emotion words between Lululemon and GymShark, we found that Lululemon-related UGC use words with stronger positive emotions while GymShark-related UGC contain words with weak positive emotion most frequently. Specifically, the plot shows that word occurrences of “love”, “hate”, “best”, and “worst” are highest in Lululemon-related UGC. Researchers have found that men often have more intense emotional experiences while women have higher emotional expressivity (Deng et al., 2016). Gender differences in emotional expressivity could be the reason for such strong emotions in Lululemon-related content across four platforms because Lululemon is the only brand whose major consumers are female.

2. Textual Similarity

To examine whether contents across four platforms are semantically or syntactically similar, we calculated the Wasserstein distance among datasets from four platforms for each brand, visualized the results as heatmaps, and transformed the level of similarity into an abstract Cartesian space using multidimensional scaling (see Figure 4). The Wasserstein distance, as known as the earth mover’s distance (EMD), is a measure of dissimilarity between two text documents “as the minimum amount of distance that the embedded words of one document need to ‘travel’ to reach the embedded words of another document” (Kusner et al., 2015).

Figure 4. Heatmaps and multidimensional scaling of Wasserstein distances across four platforms for all five brands. The lighter the color in the heatmaps, the greater Wasserstein distances found between two text datasets.



Five brands' heatmaps and multidimensional scaling plots of Wasserstein distances among four platforms demonstrate that brand-related contents are semantically different. Especially, the color of any Reddit blocks is the lightest in all five heatmaps and Reddit appears at the top right corner of all multidimensional scaling plots, indicating the Wasserstein distances between Reddit and any other platforms are the greatest. In other words, brand-related UGC on Reddit displays the most distinct pattern compared to UGC on the other three platforms. The functionalities of Reddit could explain such a distinction. Reddit is a collaboration platform where people could share their experiences using brand products, while at the same time, Reddit users also ask and seek brand-related questions and answers. Since data collected from Reddit are limited to subreddit posts, without comments and responses under each post, the majority of brand-related posts can be in the forms of questions, which thus show a distinct word probability and semantic pattern than UGC from other platforms.

Moreover, content on YouTube and Twitter are semantically more similar than those on Instagram, as we found YouTube and Twitter are close together in the multidimensional scaling plots of four brands (except Under Armour). One shared and core functional block between Twitter and YouTube is conversation. Twitter is centered around exchanging short messages that are real-time status updates, and thus considered as being more about conversation than identity, and similarly the honeycomb of YouTube (see Figure 1) also indicates conversation is one of its core functionalities (Kietzmann et al., 2011). Thus, contents on Twitter and YouTube are more casual and conversational, resulting in small Wasserstein distances between the two platforms. Besides, we found that Nike, GymShark, and Puma-related UGC on Instagram are observed to be distant from contents on YouTube and Twitter. The use of emojis in UGC is pervasive only on Instagram and the most frequent bigrams and trigrams on Instagram contain emojis across five

brands. However, the high-level textual differences measured by Wasserstein distances cannot explain how specific topics are discussed across types of social media, which are analyzed in the next step.

3. Topic Clusters

To begin with, the most frequently used bi-grams and trigrams reveal some aspects of language patterns and potential topics discussed on each platform. Specifically, most frequent bigrams and trigrams on YouTube are observed often appearing in verbal communications, like “gon, na”, “little, bit” “m, sure”, while those on Instagram are often emojis in a sequence. Bi-grams and trigrams found on Reddit indicate its contents tend to be the most product-related. Brand and product names often appear in the frequently used bigrams and trigrams on Reddit, for example, “puma, x” (with a likelihood ratio of 6.67), “r, underarmour” (4.66), “air, max” (9.7), “align, tank” (9.91), and “gym, shark” (7.38). Such a pattern is not found in UGC on other platforms. Moreover, the distinctiveness of Reddit contents is explained by the fact that consumers use Reddit as a platform to ask questions and seek answers, which is also supported by bigrams and trigrams. “Black, Friday, sale” (with a likelihood ratio of 4.24) and “black, friday, cyber” (2.82) appear to be the two most frequent trigrams in the UGC related to Under Armour and GymShark. In addition, bigrams, “thank, advance”(6.4) and “need, help” (4.39), are found to be significant in some datasets. Thus, the exploration of bi-grams and trigrams help explain the distinctive semantic patterns appearing in Reddit brand-related contents.

Next, the word clouds of all datasets (see Figure 5) confirm our previous findings on the focus of brand products and promotion on Reddit, which partially supports the second hypothesis, that UGCs on collaboration platforms are more brand-focused than those on creative outlets. From the word clouds of four brands, the brand names (i.e. "puma", "ua", "lulu" or

"lululemon", "gymshark") never appear to be the most prominent words on Instagram and YouTube, but they seem to be heavily discussed on Reddit. We also found, besides brand names, the most popular brand products are also prominent on Reddit word clouds. Specifically, "shoe", "sneaker", "shirt" are important tags in contents related to Puma and Under Armour, and "align", the most popular Lululemon leggings line, also appears to be the most prominent word in the word clouds of Lululemon-related contents on Reddit. These findings support that brands and their products are most likely to be the central topics in brand-related UGC on Reddit.

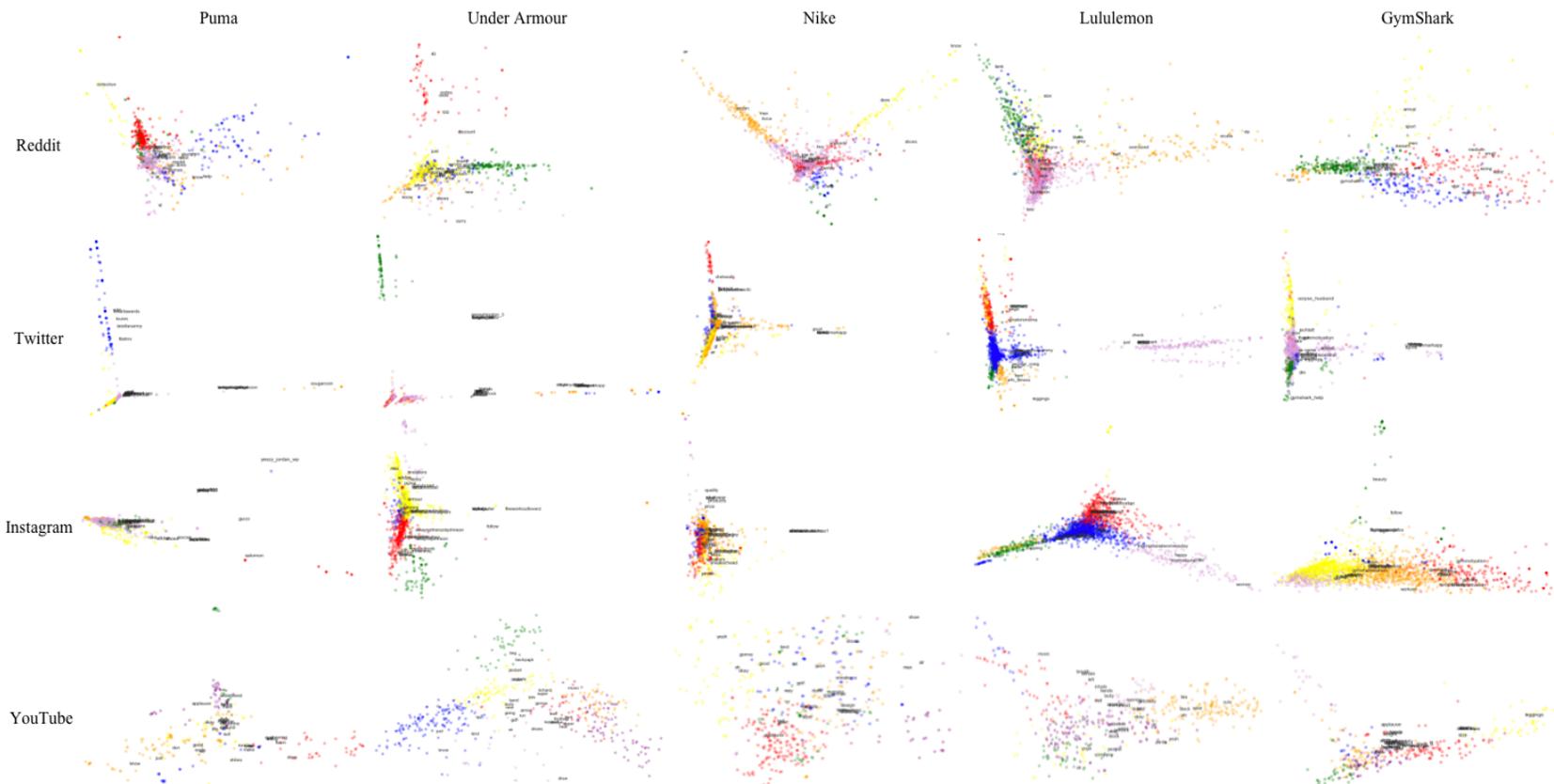
However, the same pattern was not consistent on self-media (i.e. Twitter) across all brands. Word clouds of contents on Twitter seem messy, and we cannot gather much helpful information from these, which results from Twitter users overly sharing their Poshmark listings. Poshmark is an app for selling and purchasing second-hand products, and it allows users to link their Poshmark accounts with other social media platforms to share their listings so that products listed have a higher possibility to be purchased quickly. Brand names and product info are critical contents in the Poshmark listings, and, as a result, brand-related UGC on Twitter consists primarily of Poshmark listings, making it difficult for researchers to get useful information. "Check", "listing", "share", "poshmark", "closet" make up most space in Twitter word clouds. Two brands that are most frequently shared from Poshmark are Lululemon and Under Armour. Especially after hyperlinks from Under Armour-related UGC data are removed on Twitter, its word cloud only consists of tags from Poshmark listings. One exception is Puma-related UGC on Twitter. Unlike other brand names, puma also refers to a genus in the family Felidae that contains the cougar, and, recently, cougar coin, a new cryptocurrency, caught people's attention. As a result, many tweets that discussed cougar coin used the hashtag #puma for the relevance between

those two words, and therefore, we found “cougar-coin”, “into-the-moon”, “save-cougars”, are the most prominent words in Puma-related contents on Twitter.

Compared to Reddit and Twitter, brands are no longer the main contents on Instagram and YouTube. How the brand products are used in daily life and consumers' experiences and opinions become central in brand-related UGC on Instagram and YouTube, confirming our hypothesis. On creative outlets, which satisfy consumers' self-actualizations and competence, brands are often used as symbols to construct their online images, and thus contents are found to feature consumer self-promotion by focusing on creators' personal experiences or opinions. We found words related to lifestyles, like "workout", "day", "fitness", "training", commonly appear on Instagram, and words that indicate thoughts and opinions, for example, "think", "feel", "love", "actually", are prominent on YouTube.

To discover high-level patterns of topics discussed in brand-related UGC on each type of platform, we applied a clustering method, k-means, and a two-dimensional content clustering, topic modelling, on each dataset (see Figure 6). The optimal cluster number for k-means is identified through the silhouette method. The silhouette method is a method to interpret and validate the consistency within a given number of clusters by generating a graphical representation of how well each object has been classified. By comparing the silhouette values generated for each k (ranges from 3 to 7), k with the highest silhouette value is chosen, which indicates that each object in k clusters is well-matched to its cluster and poorly matched to the neighbouring clusters. Because the best k selected for all datasets is greater than 3, we listed the top terms from 3 clusters for each dataset in a table (see Figure 6), indicating three distinct topics discussed in the brand-related UGC on each platform.

Figure 6. Visualization of k-means topic clustering of brand-related UGC on four platforms. The table below displays keywords extracted from each cluster and they indicate the topic/theme of one cluster.



Platform	Brand	Cluster	Keywords	Topic/Theme
Reddit	Puma	Cluster 1	help, identify, need, shoes, shoe, id, identifying, pls, puma, pair	help, identify, identifying, shoes, need, thanks, finding, appreciated
		Cluster 2	collection, puma, rate, new, addition, added, latest, wheel, hoops	know, does, shoes, new, buy, want, model, shoe, pair, got
		Cluster 3	shoes, know, does, puma, model, buy, love, new, wearing, sneakers	ua, just, shirt, like, shorts, size, shirts, project, ve, compression
Twitter	Under Armour	Cluster 1	cougarcoin, cougarinthemoon, cougaringlobal, betterhandoge, savecougars, cougartr, cougar, ilonmask, trustwalletapp	40, codes, code, discount, promo, pn, thanks, coupon, ambassador
		Cluster 2	poshmark, loving, share, fashion, items, check, shopmycloset, style	loving, share, poshmarkapp, items, good, check, poshmark, fashion, angelposhshop
		Cluster 3	pumafootball, like, new, just, don, adidas, kits, bds, think, people	new, armour, ua, amp, shop, season, therock, men, time
Instagram	Nike	Cluster 1	nike, sneakers, adidas, jordan, shoes, yeezy, fashion, style, available	run, running, runnersofinstagram, runningmotivation, runner, garmin, fitness, uarunning, miles, asics
		Cluster 2	football, soccer, pumafootball, neymar, manciny, premierleague, fifa, psg, nike	run, running, runnersofinstagram, runningmotivation, runner, garmin, fitness, uarunning, miles, asics
		Cluster 3	nike, adidas, instagood, love, fashion, model, new, follow, day	nike, adidas, armour, puma, new, reebok, day, fashion, size, women
YouTube	Lululemon	Cluster 1	shoe, foam, soft, cushioning, lining, features, lace, design, leather	shoe, run, shoes, golf, kind, running, know, speed, think
		Cluster 2	casadora, snow, life, ground, rule, moment, la, fromthe, beautiful, expect	lots, super, support, comfort, ua, calm, foot, giving design, flexible, tongue, synthetic, lace
		Cluster 3	know, just, don, really, good, want, got, right, think, yeah	bag, backpack, pockets, main, shoulder, soft, padded, nice, space
YouTube	GymShark	Cluster 1	help, identify, need, shoes, shoe, id, identifying, pls, puma, pair	help, identify, identifying, shoes, need, thanks, finding, appreciated
		Cluster 2	collection, puma, rate, new, addition, added, latest, wheel, hoops	know, does, shoes, anybody, buy, called, model, shoe, kind, thanks
		Cluster 3	shoes, know, does, puma, model, buy, love, new, wearing, sneakers	air, max, force, jordan, nike, 90, 1s, custom, pair, high

Starting with Twitter, the topic clusters also reveal the overwhelming Poshmark listings on the site. "Poshmark, shop-my-closet, check, share, listings" is a common cluster discussed in UGC related to all five brands. Besides, other brand-focused clusters are also found in clusters of Twitter contents. For example, because Under Armour has announced a partnership with Dwayne "The Rock" Johnson, an actor and former professional wrestler, "armour, the-rock, project-rock" is one of the topic clusters discussed on Twitter. Similarly, Chelsea Football Club has announced a partnership with Nike, and "new, chelsea, chelseafc, kit" is one of the clusters in Nike-related contents. Interestingly, brand partnership is one of brand-related topics found only on Twitter. For a brand that is not focused on partnership (i.e. GymShark), complaining might be one of the motivations of consumers' tweets. Customer service, or service responses have been found as one of the motives that drive brand-consumer interactions on social media (Rohm et al., 2013; Hamilton et al., 2016). "Gymshark_help, dm, support, help, new, order, fix" and "thank, love, appreciate, customer" are found as two topic clusters in GymShark-related tweets, which confirms findings from previous studies, that consumers' primary motivation for complaining on Twitter is to ask firms to resolve issues" (Ma, Sun, and Kekre, 2015). Thus, although Poshmark lists make it difficult to find patterns of UGC on Twitter, topic clusters still help discover some brand-related topics, like brand partnerships and communications between customers and brands to solve issues with their orders or products.

Clusters of UGC on Reddit affirms that Reddit, as a collaboration platform, is used by consumers to ask questions about brand products, and thus, UGC on Reddit are brand-centered and product-focused. This pattern is consistent across the five brands studied. "Help, identify, pls, need, thanks, find, appreciated" is a topic cluster that appears in both Puma and Nike related contents. Brand signature products and brand names, for example, "puma, shoes, model, buy,

sneakers", "air, max, force, nike, custom, pair", and "lululemon, leggings, bras, align tank", always appear together in one central topic cluster of Reddit brand-related UGC. Moreover, another brand-centered topic, promotion, appears in multiple clusters of Reddit contents. "Code, discount, promo, coupon, ambassador" and "sale, gymshark, going, website, code" are two clusters that appeared in Under Armour-related and GymShark-related contents, respectively. The discussion of brand promotion and sales is only found in UGC on Reddit. Thus, Reddit contains the most brand-centered discussions compared to the other three platforms, which is not surprising. Posts are organized by subject into user-created boards, known as subreddits, and the discussion under each subreddit is generated around the central topic, which is the brand.

Different from self-media and collaboration outlets, UGC on creative outlets focuses on experiences, lifestyles, and opinions around brand products, whereas YouTube content is more information-focused while Instagram content is more self-promoting and motivation-focused. Our hypothesis, that content on creative outlets, YouTube and Instagram, focuses on its creator's personal experiences or daily life, is not fully supported by the results. In fact, many topic clusters of YouTube content also contain brand products, like shoes, sports bras, leggings, and shirts, but topics discussed are not simply about brand products. For example, the cluster that contains the product "shoe" in Puma-related UGC on Reddit is "shoe, know, does, puma, model, love, buy", while the cluster that contains the same word "shoe" in Puma-related UGC on YouTube is "shoe, foam, soft, cushioning, lining, feature, design". Although the two clusters are both about Puma's "shoe" products, Reddit users try to ask about a specific shoe model made by Puma (brand-centered), but YouTube users discuss features and designs of sports shoes in general (not brand-centered). Similarly, as Nike-related UGC on Reddit focuses on "air, max, force, shoe, nike, custom", Nike-related UGC on YouTube is about "shoe, runs, collar, outsole,

design, fabric, foam". Besides, brand names do not appear in any YouTube topic clusters with those brand products, but the activities in which those products are used accompanies the discussion of products. "Shoe" and "running" appear in the same cluster, and "leggings" and "workout" or "yoga" appear in the same cluster. Interestingly, one topic cluster of GymShark contains "business, market, ben, francis, growth, billion, company". Ben Francis is the founder of GymShark, and it shows the reasons behind this fast-growing company and its founder are also topics YouTubers like to discuss. As a result, topic clusters of UGC on Twitter and YouTube confirms what Smith et al. (2012) found in their study, that brands are more likely to be central in brand-related UGC on Twitter and less likely on YouTube. However, Smith et al. also found that brand-related UGC on YouTube are the most self-promoting and thus the least brand-central, which is not entirely accurate in this study. Although YouTube, as a creative outlet, engages creators to be the stars of their videos and its culture supports the development of micro-celebrities (Burgess and Green, 2009), brand-related UGC on YouTube extends beyond basic brand discussion and provides information related to a consumer product or product-related activity, rather than being solely self promotion of content creators.

However, consumers on Instagram are shown to promote themselves through brand-referencing communications. Most of the topic clusters of Instagram UGC are not about the brand or even their products, but they are centered around posters' personal lifestyles, experiences, goals, and attitudes. Since the industry chosen is athletic apparel, "workout, fitness, gym, training, bodybuilding, motivation" is one common cluster that appeared across many brands, except for Puma. In fact, the exception of Puma might be a result of its marketing strategies and brand image. In March, Puma just released its long-term partnership with Neymar, a famous Brazilian football star who had ended his previous partnership with Nike. Thus,

Puma-related contents are related to the topic "football, soccer, neymar, mancinity, premier-league, fifa, psg, nike". In addition, compared to other brands, Nike, the leading athletic footwear brand, is also considered to play an essential role in the fashion industry. Two unique topic clusters are found only in Nike-related UGC on Instagram, "photo-of-the-day, love, photography, fashion, model, style, follow, ootd" and "sneakers, kickstagram, hypebeast, sneaker-addict, sneakerheads, am1-collector, kiss-my-airs", indicating Nike's two consumer segments, fashion-lovers and sneaker-lovers. #fashion is also the most used hashtag when Instagram users make #nike posts.

Moreover, patterns of Instagram content confirm that altruism and social identification are the two primary motivations for consumers to engage in brand communities (Lee, Kim, & Kim, 2011). "Women, day, international, happy, strong, world, women-supporting-women" is one topic cluster found in Lululemon-related contents, and "gymshark-women, happy, gym, beauty, fitgirl" is also a related topic cluster found in GymShark-related contents. Creating brand communities allows customers to share their attitudes, identities and encourage others by brand-referencing communications. Contents related to GymShark and Lululemon are also most motivation-focused and workout-related, showing that consumers in brand communities like to share contents that are encouraging or helpful to other members in the community. Besides, Instagram topic clusters also support that engagement in brand communities help consumers develop trust and loyalty towards the brand (Habibi et al., 2016; Coelho et al., 2018). We found a tendency in Instagram users to include brand competitors in their contents, yet this pattern is only found in brands that lack strong brand communities (i.e. Puma, Under Armour, and Nike). First, Instagram users are known for using many hashtags in their contents to get attention, and thus, tagging brand competitors or other relevant brands is a typical behaviour only on Instagram. Adidas, another leading athletic brand (not included in this study), appears in the topic clusters of

contents related to Nike, Puma, and Under Armour. Clusters generated by Puma and Under Armour-related UGC also contain Nike. Mentioning of brand competitors is not found on other platforms or brands with a strong sense of community.

In conclusion, to find out what topics consumers like to discuss on each platform, we constructed the word clouds of each dataset and conducted k-means clustering and topic modellings. The results support our second hypothesis with some unexpected variations. UGCs on self-media and collaboration platforms are indeed more brand-focused than those on creative outlets. In particular, consumers use Reddit as a platform to ask product-related questions and seek promotions. Nevertheless, Twitter users are overly sharing their Poshmark listings, making Twitter less brand-focused, but the discussion of brand partnership and news is prominent on Twitter. Brands are peripheral topics on creative outlets. Content on YouTube is found to be most informative, by discussing a general product or focusing on the product-related activities, like running or yoga. Brand-referring contents on Instagram are heavily focused on creators' self-presentation and self-promotion. The sense of brand community plays a vital role in creating contents on Instagram, leading to more motivation-focused and beneficial content found.

4. Topic Popularity

To explore what topics discussed in brand-related UGC are more likely to receive attention on each type of social media, we first ran LDA and generated ten topics for each dataset, and then visualized topic distributions of the 10 most popular posts (highest attention) and 10 least popular posts (lowest attention) (see Figure 7-11). LDA identifies each topic with a different probability of generating each word in the document. After we ran LDA on all datasets with different topic numbers, we found 5 or 6 as the smallest number of topics to produce insightful words within each topic.

Figure 7. Topic Distribution of Most (left) and Least (right) Popular Puma-related UGC

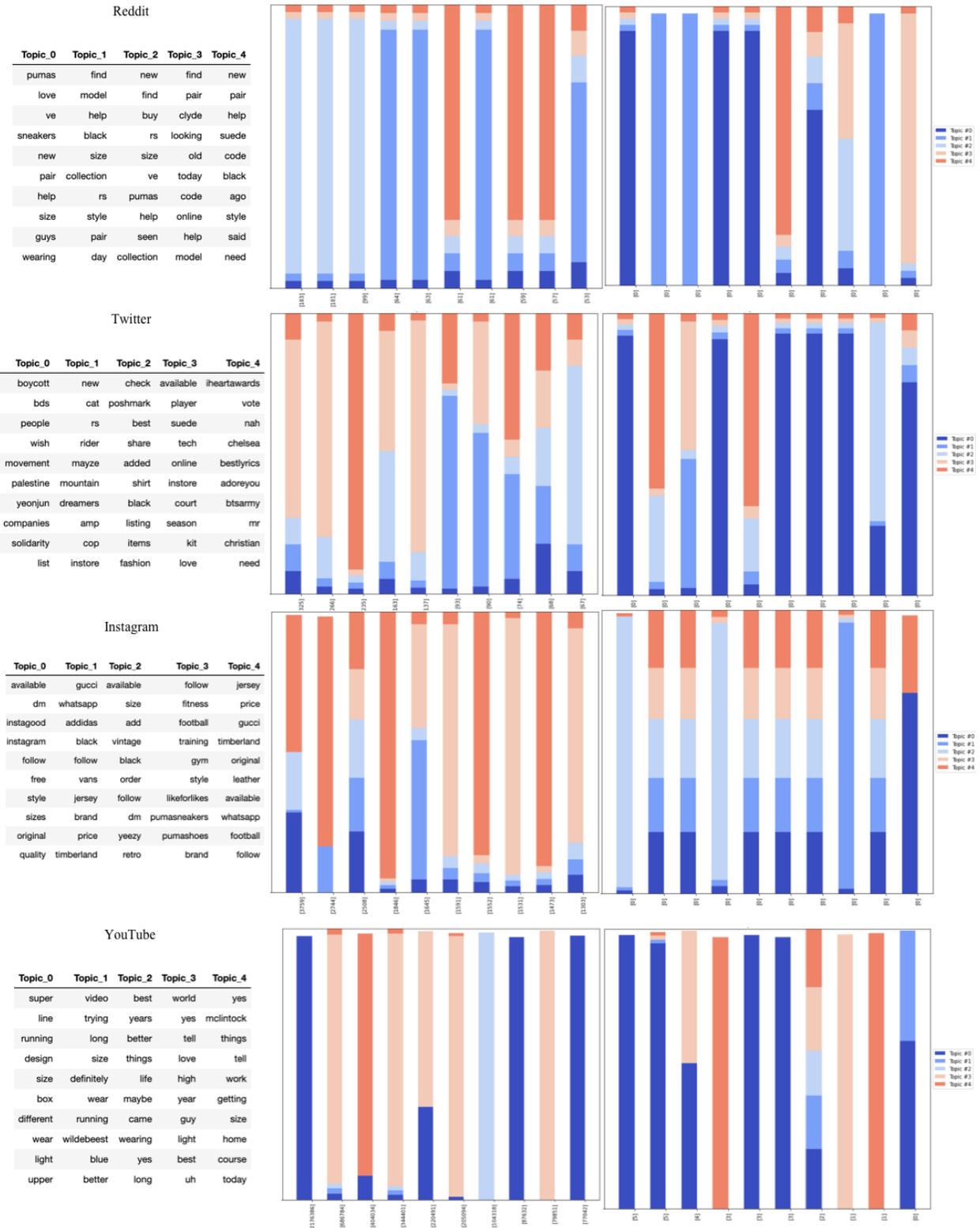


Figure 8. Topic Distribution of Most (left) and Least (right) Popular UA-related UGC

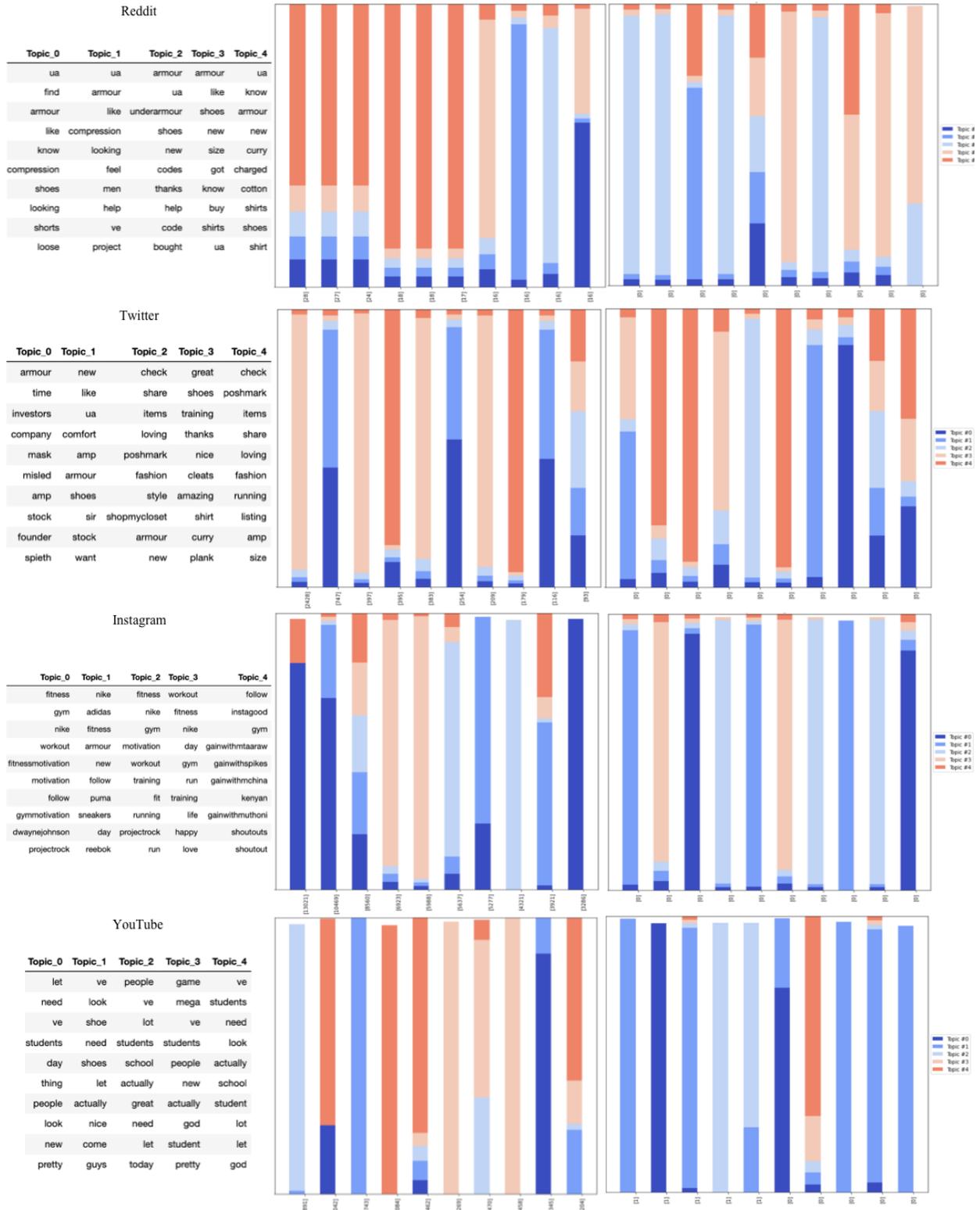


Figure 9. Topic Distribution of Most (left) and Least (right) Popular Nike-related UGC

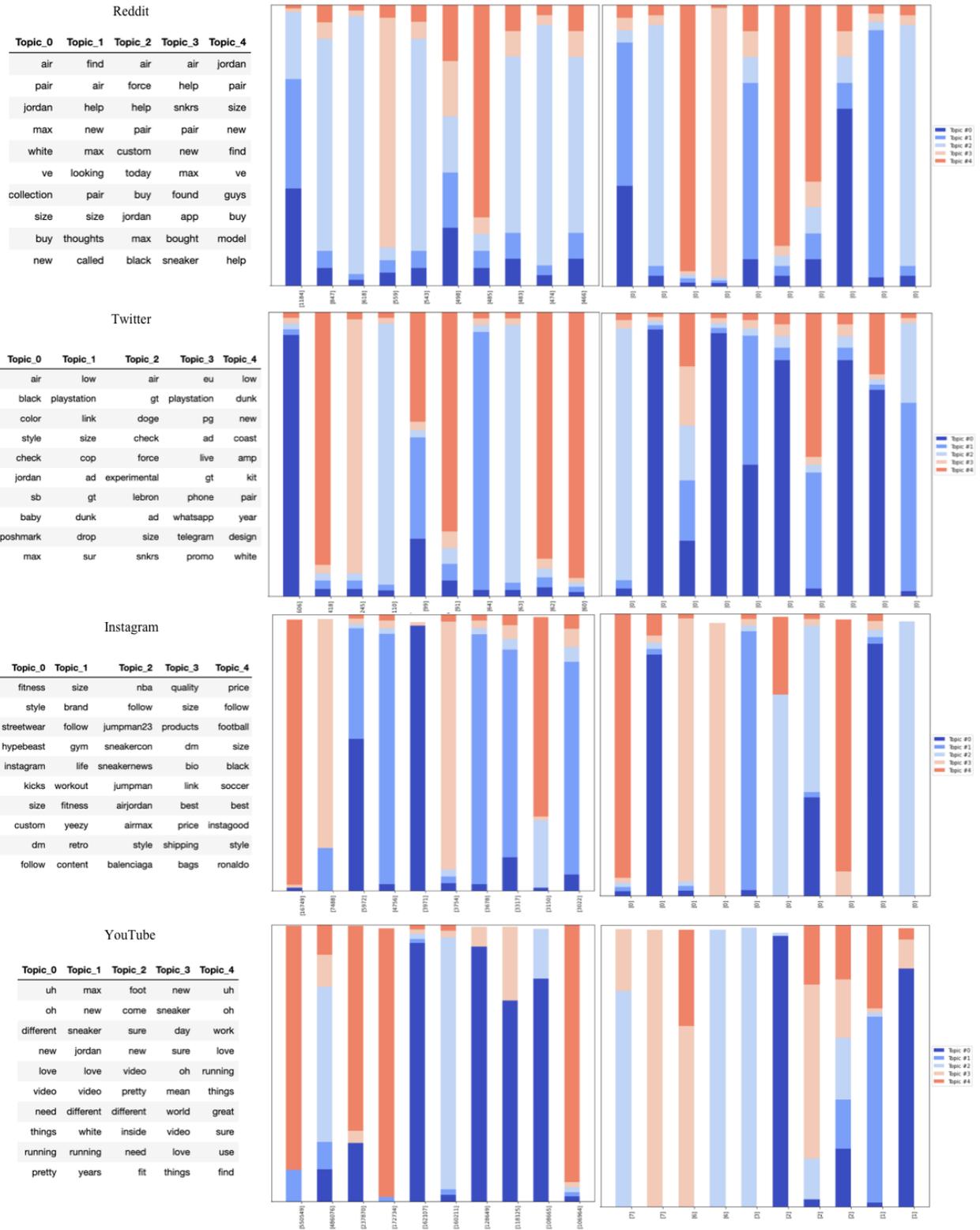


Figure 10. Topic Distribution of Most (left) and Least (right) Popular Lulemon-related UGC

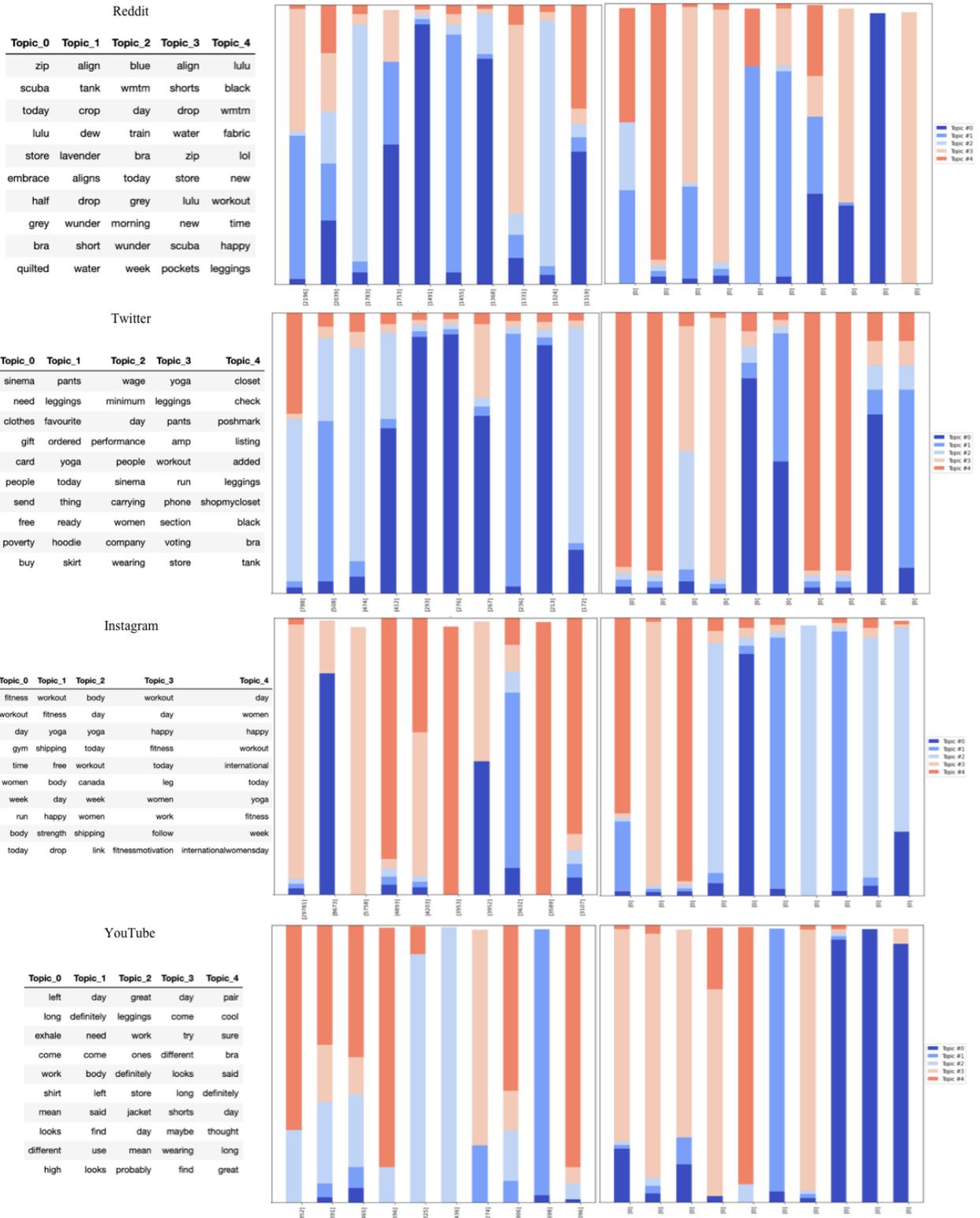
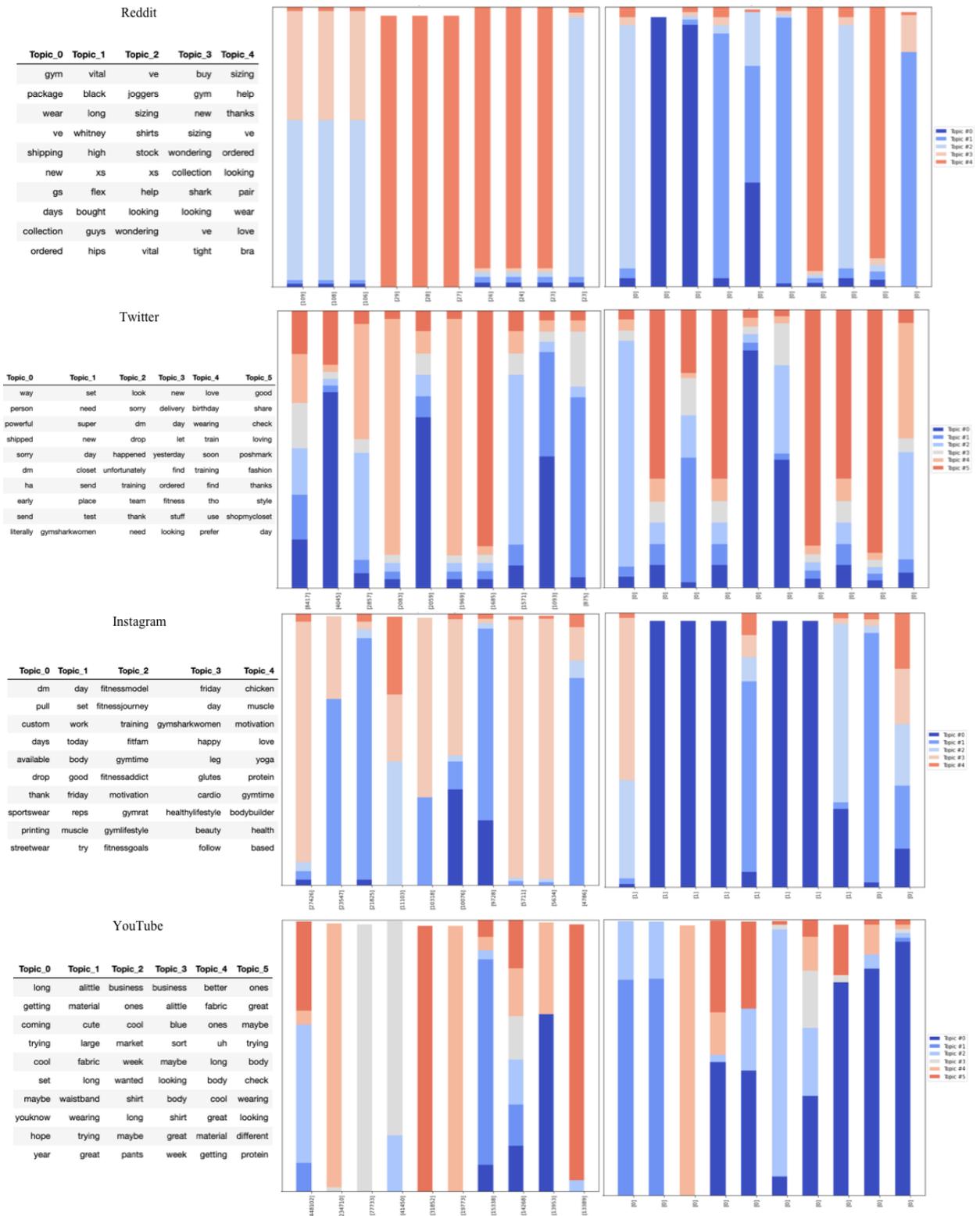


Figure 11. Topic Distribution of Most (left) and Least (right) Popular GymShark-related UGC



As a result, each identified topic contains a list of top words and associated coefficients that indicate the probability of each word belonging to the topic. However, we found frequently-used words appearing in multiple topics, which affect our interpretations, especially in YouTube scripts. Thus, we removed 50-100 of the most frequent words to purify YouTube data so that results from LDA could better reflect the actual topics discussed by consumers. Then, we examined the percentage of each topic's presence in one post. The higher percentage one topic takes in a post, the closer its content is around that single topic.

The comparison between topic distributions of the 10 posts that received the most attention and those of 10 posts that received the least attention in the dataset reveals interesting patterns of what contents draw users attention to different types of social media. Starting with Reddit, asking for help to find a particular model or asking for suggestions on products or sizes are found to be what is mainly discussed on Reddit. Autonomy, one of the 10 human needs (Sheldon et al., 2001), is the main need satisfied by viewing contents on collaboration platforms. Users voluntarily browse Reddit posts to learn news things and look for answers, and thus, we found questions related to a brand's popular products receive more attention than other general questions. Specifically, the top posts are found to directly relate to the Under Armour Curry Brand shoes made with charged cotton, Nike's customizable Air Force brand, and Lululemon Align and Wunder lines. Other topics about consumer products, like shoes, shirts, bras, without mentioning specific model or brand product lines, receive less attention on Reddit.

However, although the popularity of UGC containing questions or asking for help depends on what products Reddit users asked about, the one exception among the five brands is GymShark, on which all question-related posts receive much more attention than posts without questions. Topic 3 (buy, new, sizing, wondering, collection, looking) and topic 4 (sizing, help,

thanks, ordered, looking, pair, wear, love) are found to be the main topics top 10 posts are about, while 10 least popular posts are about topic 0 (package, wear, shipping, new, days, collection) and topic 1 (vital, black, long, whitney, flex, bought). This result supports that consumers in well-formed brand communities would like to help others more because of altruistic motivations.

On Twitter, the Poshmark listing is one of the topics appearing across five brands. The topic distributions of tweets with the highest number of retweets and likes show that nearly none of them are related to Poshmark and selling products, specifically topic 2 in UA-related UGC (check, share, items, loving, poshmark, fashion), topic 0 in Nike-related UGC (air, black, color, check, style, poshmark), topic 2 in Puma-related UGC (check, poshmark, best, share, added), topic 4 in Lululemon-related UGC (closet, check, poshmark, listing, added), and topic 5 in GymShark-related UGC (good, share, check, loving, poshmark, fashion). The topics that receive the most attention on Twitter are brand-dependent. Company news, brands' newly released products, and new brand collaborations commonly appear in the most popular brand-related tweets, for example, Under Armour's stock and the company's founder (in topic 0 and 1), Nike's upcoming PlayStation (in topic 1 and 3), Dunk Lows (in topic 4), and LeBron shoes (in topic 2).

Moreover, as a micro-blogging form of self-media, Twitter is where users share and report information, opinions, daily conversations, news, and complaints (Java et al., 2009). Brand-focused news, including company's news, brands' newly released products, and new brand collaborations, are observed as topics in Nike and UA-related UGC. Whereas the most discussed topics in Lululemon and Puma-related tweets are recent news, brands are found to be peripheral in the news. Krysten Sinema, who voted against a \$15 living wage but was found carrying a Lululemon backpack, becomes the subject of Lululemon-related topics in the most popular tweets (see in topic 0 and 2). BTS, a seven-member South Korean boy band, who recently

became Puma's global ambassador, won the iHeartRadio Music Awards for their new song, which then was found one of the popular topics in Puma-related tweets (see in topic 4). In addition, Puma claimed their support on the Israeli occupation, leading to a boycott movement hold by the Palestinian BDS National Committee, which is found to be one of the topics in Puma-related tweets (see in topic 0) but did not receive much attention. However, consumer complaints and marketers' responses are found to be the main topics in popular tweets related to GymShark (see in topic 0 and 2), rather than news and trends. In conclusion, examining popular and unpopular topics in brand-related tweets revealed consumers' interests in current brand-related news and trends, yet brands may not always be the subject of matter.

Although, in the previous section, we found brand-referring contents on Instagram are heavily focused on creators' self-presentation and self-promotion, since last year, Instagram released a new shop function, which allows business accounts to create shops and users can place orders directly through Instagram. This function makes Instagram no longer a platform for only creators but also for users who want to start their businesses. Thus, we found product-focused content or posts to promote sales in Instagram posts related to all five brands. However, brand-related UGC receives the most popularity when Instagram creators talk about their lifestyles and motivations rather than selling or listing products. Pleasure or simulation is the main need that is satisfied by browsing contents on creative outlets (see Table 1; Zhu and Chen, 2015). Thus, UGC on Instagram that provides pleasure and simulation for viewers received more popularity than product-selling posts. For example, topic 0 (fitness, gym, workout, fitness-motivation, gym-motivation) and topic 3 (workout, fitness, day, run, training, life, happy) are the main two topics in the top 10 most popular Under Armour-related posts. Topic 0 (fitness, style, streetwear, hypebeast) and topic 1 (gym, life, fitness, workout) also appear more frequently

in the top 10 most popular Nike-related posts. Puma-related UGC that discussed topic 3 (fitness, football, training, gym) received more likes. Similarly, topics related to body positivity, fitness, and female power, receive the most attention in Lululemon and GymShark-related contents (topic 3 and 4 in Lululemon-related UGC and topic 1 and 3 in GymShark-related UGC), compared to topics related to selling brand products. Specifically, "free, shipping, drop, dm, available" appears in Lululemon-related topic 1 and 2 and GymShark-related topic 0, which are topics found in the least popular posts. The comparison among five brands shows that contents that are related to brands with a strong sense of communities (i.e. Lululemon and GymShark) are more likely to be motivation-based and lifestyle-focused, and product-based content only takes a small portion of posts and is not likely to be liked by other users. However, brands that focus more on promoting their products than creating online communities (i.e. Puma and Nike) contain a mix of self-promoting motivation-focused contents and product-focused contents on Instagram.

Unfortunately, although we have filtered out 50 to 100 frequently used words in YouTube scripts to generate insightful topics, it's still difficult to infer each topic based on words given because they are used widely in spoken languages, for example, "actually", "definitely", "pretty", "need", "try", "let", etc. Besides, because videos are made in different lengths and scripts vary in length too, one extremely long video over a single topic can result in words from that video being overweight. For example, Under Armour planned and produced the All-American Football Game in 2020 and, due to COVID, hours-long videos recorded are shared on YouTube, watched and liked by many audiences, resulting in words like "student", "school", and "game" being repeated in multiple topics and become the most popular topics. However, we still have a few interesting findings. First, none of the topics are found to be related to particular brand products or product lines, except topic 1 in Nike-related UGC (max, sneakers, jordan),

which is not yet discussed in any popular videos. Besides, most activity-related topics like running and working out are found in the popular videos. Lastly, although particular brand products are not commonly found in brand-related YouTube videos, the discussion of general products is pervasive, from different aspects (i.e. design, material, and fabric).

As a result, to examine what topics are more likely to receive attention from other users, we conducted LDA topic modelling on each dataset and visualized the topic distribution of the top 10 most and least popular posts. Our third hypothesis (the most popular content is opinion-based on self-media, product-focused on collaboration platforms, and experience-based on creative outlets) is not supported. The most popular contents on self-media (i.e. Twitter) are found to be trends/news-focused, and brands are often the centre of news topics with a few exceptions. As we found earlier, consumers use collaboration platforms (i.e. Reddit) to ask product-related questions and seek promotions. The most popular topics are found to be question-based, especially for brands with strong brand communities. Although some creators promote brand products on creative outlets, in large part those product-related topics did not receive significant attention. Contents related to creators' lifestyles, motivations, and achievements are more likely to become popular on Instagram, especially for community-focused brands. Activity-related content or reviews of consumer products are popular topics on YouTube.

CHAPTER 5

DISCUSSION

1. Benefits

To explore how channel characteristics would influence consumer creations of brand-related messages and what content receives more attention, we extracted UGC from four platforms (Twitter, Instagram, YouTube, and Reddit) that characterize different types of social media. Five athletic wear brands are chosen for this study: Nike, Puma, Under Armour, Lululemon and GymShark, which share different consumer profiles and deploy different marketing strategies. We deployed machine learning methods to derive brand topics and classify brand sentiments from big UGC data, majorly through sentiment analysis, k-means clustering, and LDA topic modelling. With more than 50,000 UGC collected from four social media platforms, this study answered what motivations drive consumers to post brand-related UGC and what content characteristics are found on cross types of social media, which are valuable to both researchers and brand managers.

All three hypotheses we proposed are partially supported by findings, with some exceptions and further interpretations. First, the comparison of brand sentiments across types of social media demonstrates that there are more positive brand sentiments observed on content-based platforms (i.e. Instagram, YouTube, and Reddit) because they satisfy user needs for self-actualization and competence. Twitter emphasizes forming conversations between users as a profile-based platform, which explains why brand-related UGC contain primarily neutral or slightly negative sentiments on Twitter. Besides, users on content-based platforms use more words with strong emotions (love, hate, best) than users on profile-based platforms.

Second, the comparison between topics found in brand-related UGC across platforms supports our second hypothesis that UGC on self-media and collaboration platforms are brand-centered. Brand-centered topics found on those two types of social media include product-related information (sizing, model, etc.), brand news and partnership, and customer services. Whereas UGC on Instagram is the most self-promoting through discussing creators' lifestyles and motivations, UGC on YouTube evaluates consumer products without focusing on their brands and introduces activities related to those products, like working out and yoga.

Finally, although brand-centered UGC is commonly found on self-media and collaboration platforms, the most popular content on Twitter is related to brand news and trends, and posts regarding the most popular brand products or product lines are found to receive the greatest popularity on Reddit. As we expected, brand-centered topics, like brand products, news, and promotions, did not receive much attention on creative outlets, because viewers enjoy contents that provide them with pleasure and entertainment on creative outlets. Contents related to creators' lifestyles, motivations, and achievements are more likely to become popular on Instagram, and activity-related contents or reviews of consumer products are popular topics on YouTube.

In addition, brand communities play an essential role in cross-platform content patterns. Consumers of brands that have formed a strong brand community are found to express more positive attitudes towards the brand across social media platforms, implying the formation of consumer loyalty and trust. Brand communities create the most influence on Instagram, compared to other platforms. Contents related to GymShark and Lululemon are found to be highly self-promotional and motivational, and those are least likely to

mention brand competitors. Because consumer engagement in brand communities comes from intrinsic motives of altruism and social identification, GymShark and Lululemon's consumers tend to share contents about their identities and contents that are encouraging and helpful to other community members.

Prior literature divides social platforms into four types based on their functionalities, user experiences, and fulfilment of distinct psychological needs. Thus, four social media platforms in this study belong to three groupings. Twitter, as a self-media platform, focuses on conversation and satisfies users' needs for popularity and money, which helps explain why many users share their Poshmark listings on Twitter. Additionally, Twitter is where users engage in and spread information, opinions, daily conversations, news, and complaints, and thus are the least likely to feature consumer self-promotion. Brand-related news, such as newly released products and company news, is discussed in the most popular brand-related tweets. For proactively managed brands (i.e. GymShark), Twitter can be an excellent channel for customers to communicate with brand managers and marketers, which help build positive brand sentiment. Besides, the negative brand-related news and complaints that consumers extensively discuss on Twitter require brand managers to respond urgently and adequately to improve their brand images and customer satisfaction. The significant amount of discussion in brand-related news also indicates Twitter, or self-media, can be an excellent way to release a company's news or new products and engage consumers to discuss.

Instagram and YouTube, as creative outlets, satisfy creators' needs for competence and self-actualization and viewers' needs for pleasure. Therefore, the inherent characteristics of creative outlets make them a good place for brands to promote

themselves and create a strong sense of brand communities. Although UGC on creative outlets is not product-focused and brand-centric, the discussion of consumers' lifestyles, values, and positive attitudes towards brands on Instagram are found to be more valuable. Similarly, on YouTube, brands appear to play a supporting rather than central role in consumers' self-presentation, which does not mean UGC on YouTube is devoid of brand information (Smith et al., 2012). Factual information in brand-related YouTube videos (for example, the design, fabric, functions of products and activities in which brand products can be used) may not promote the brand itself but help place the brand in consumers' consideration sets. Thus, YouTube can be a great platform for people who are not looking for particular brand products but learn about general fitness products. The characteristics of creative outlets imply opportunities for brand managers to collaborate with consumers to circulate positive sentiments about brands and increase the visibility of brands.

Lastly, as a collaboration platform, Reddit is built on groups and collaborative conversation and satisfies users' autonomy in brand-related information search. The primary focus on product-related questions and brand promotions makes Reddit a good channel for consumers who are looking for information about particular brand products. Marketers can also use Reddit as a platform to answer consumers' questions about products to increase the likelihood of purchasing. Especially, posts related to popular brand products receive more attention indicating consumers' interests in the brand products.

User-generated contents (UGC) have significant potential business value in targeted advertising, brand communication, and customer engagement (Liu et al., 2017). This study integrates the previous framework and expands beyond the limitations of previous literature: single-platform focus (Liu et al., 2017) and limited dataset (Smith et

al., 2012), by using actual big data related to multiple brands of the athletic retail industry on four social media platforms. Secondly, instead of being platform-based, this study compares theoretical and conceptual categorizations of social media platforms proposed by prior literature and analyses the content patterns based on users' psychological needs satisfied by posting and browsing on each type of social media. When new social media sites are created, the results from this study can still shed light on UGC on those platforms according to their channel characteristics. Lastly, this study can provide insight on future user-brand communications, because it reveals what characteristics of social media platforms can bolster the communication of distinct brand objectives with different target groups. Our results also suggest that brands should take a proactive social marketing strategy by creating and reinforcing brand communities where conversations are formed and led by consumers.

2. Limitations

Although this study has confirmed and extended insights in previous studies (Smith et al., 2012; Liu et al., 2017), by broadening the scope of the study, utilizing big data and machine learning methods, and analyzing based on channel characteristics and users psychology, there are some existing limitations that we acknowledge in our methodology. First, due to the limitations of APIs and the limited number of UGC on specific platforms, we collected data in different sample sizes for each platform, which affected variations on topics found within each dataset. However, the analysis of five brands can mitigate differences in sample sizes cross-platform and still affirm that results can be extrapolated beyond a single brand.

Second, because each UGC is posted at a different time, its chronological order can influence how much attention it had received at the moment when we scraped the post. As a result, UGC posted earlier are more likely to receive more views than UGC posted lately, and some unpopular posts we studied may receive many likes and upvotes in the near future. However, because our goal is to examine what topics are more likely to be popular, the popularity of UGC that received the highest number of likes and upvotes is certain, and thus topics found in those posts are still valuable to answer our research question.

Third, as prior studies have found, sponsored content on social media received less favorable reactions from users and can induce their cognitive and affective resistance (Kim and Song, 2018; Van Reigersdal et al., 2016). However, in this study, due to the limitation of APIs and scraping methods we deployed, we did not filter out the sponsored content in our analysis. Besides sponsored Instagram posts, tweets, and YouTube videos, many brands (Nike, Under Armour, Lululemon, and GymShark) have started the brand ambassador program to promote their brands through UGC on social media created by influencers and content creators. Although those UGC work similarly as sponsored contents, they are more difficult to be identified when we collect UGC data. The number of ambassadors and sponsored content created by each brand may influence our results. For example, GymShark promotes its brand through building an online brand community and inviting influencers and creators to become GymShark athletes, and thus top popular contents that focus on fitness and motivations can be produced by its brand ambassadors.

Next, because topic words we found through LDA on YouTube video scripts appear commonly in spoken language and daily dialogues, it is challenging to determine brand-related information discussed in those videos. So forth, results from LDA and

sentiment analysis need more effective and accurate means of validation for their findings (Liu et al., 2017), and current automatic text analysis requires extensive and problem-specific validation, which can only be achieved by careful thought and close reading (Grimmer and Steward, 2013).

Lastly, although our study focused on different types of brands to ensure content patterns can be extrapolated (variations in companies' sizes, marketing strategies, brand images, target consumers), we only focused on one industry to study, and thus, our findings may only apply to athletic wear.

3. Future Research

Our implementation of machine learning methods on big UGC data to generate brand insights on multiple social media ascertains the framework deployed by Liu et al. (2017), which can be helpful for future research works. Although our study tried to touch on different types of social media, we only extracted UGC from one or two platforms within each categorization. Future studies could analyze additional sites and platforms within every four types of social media, for example, Pinterest as another creative outlet and Quora as another collaboration platform, and brand insights and implications can be more generalizable. Dimensions of content characteristics can also be extended by future research. Because we only focused on brand sentiments and topics and examined popular topics found on each platform, future researchers could focus on the interactions between marketers and consumers, how popular brand-related UGC spreads to large audiences and goes viral, and other social media metrics that are helpful for brand marketers (Hoffman and Fodor, 2010). Additionally they could extend the study to image analysis as many UGCs are in imagery form. Sponsored and non-sponsored UGC is another direction to

compare and explore, as we have mentioned the limitations existing in this study.

Increasing characteristics of UGC studies can enhance our understanding of how each type of social media plays a role in generating those contents. Lastly, future researchers can also apply our framework to a broader representation of industries or a wider variety of product/service categories. Although our study has revealed some influence of brand communities in consumers' engagement in generating brand-related UGC, future research could also expand on how brand communities influence consumer use of social media. Nonetheless, through text analysis by machine learning models, this study illustrated practical implications in how channel characteristics can influence characteristics of brand-related UGC, providing insights for future research and brand managers.

REFERENCES

- Alhabash, Saleem, Juan Mundel, and Syed Ali Hussain (2017), "Social Media Advertising: Unraveling the Mystery Box," in *Digital Advertising: Theory and Research*, Shelly Rodgers and Esther Thorson, eds., New York: Routledge, 285–99.
- Berger, J., & Iyengar, R. (2013). Communication channels and word of mouth: How the medium shapes the message. *Journal of Consumer Research*, 40(3), 567–579.
<https://doi.org/10.1086/671345>
- Berthon, P. R., Pitt, L. F., Plangger, K., & Shapiro, D. (2012). Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy. *Business horizons*, 55(3), 261-271.
- Boerman, Sophie & Reijmersdal, Eva & Neijens, Peter. (2014). Effects of Sponsorship Disclosure Timing on the Processing of Sponsored Content: A Study on the Effectiveness of European Disclosure Regulations. *Psychology & Marketing*. 31. 10.1002/mar.20688.
- Burgess, Jean and Joshua Green (2009), *YouTube (Digital Media and Society Series)*. Cambridge, UK: Polity Press.
- Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3), 345–354.
<https://doi.org/10.1509/jmkr.43.3.345>
- Christodoulides, George & Jevons, Colin & Bonhomme, Jennifer. (2012). Memo to Marketers: Quantitative Evidence for Change How User-Generated Content Really Affects Brands. *Journal of Advertising Research*. 52. 10.2501/JAR-52-1-053-064.

- Coelho, Pedro & Rita, Paulo & Santos, Zelia. (2018). On the relationship between consumer-brand identification, brand community, and brand loyalty. *Journal of Retailing and Consumer Services*. 43. 10.1016/j.jretconser.2018.03.011.
- Daugherty, T., Eastin, M. S., & Bright, L. (2008). Exploring consumer motivations for creating user-generated content. *Journal of interactive advertising*, 8(2), 16-25.
- Deng, Y., Chang, L., Yang, M., Huo, M., & Zhou, R. (2016). Gender Differences in Emotional Response: Inconsistency between Experience and Expressivity. *PloS one*, 11(6), e0158666. <https://doi.org/10.1371/journal.pone.0158666>
- Dron, J., Designing the undesignable: social software and control. *Educational Technology & Society*. 10(3). 60- 71. 2007.
- Gensler, S., Völckner, F., Liu-Thompkins, Y., & Wiertz, C. (2013). Managing brands in the social media environment. *Journal of interactive marketing*, 27(4), 242-256.
- Goh, K. Y., Heng, C. S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information systems research*, 24(1), 88-107.
- Gómez, M., Lopez, C., & Molina, A. (2019). An integrated model of social media brand engagement. *Computers in Human Behavior*, 96, 196-206.
- Grimmer, Justin, and Brandon M. Stewart (2013), "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," *Political Analysis*, 21, 267–97.
- Habibi, Mohammad Reza & Laroche, Michel & Richard, Marie-Odile. (2016). Testing an extended model of consumer behavior in the context of social media-based brand communities. *Computers in Human Behavior*. 62. 292-302. 10.1016/j.chb.2016.03.079.

- Hamilton, M., Kaltcheva, V. D., & Rohm, A. J. (2016). Hashtags and handshakes: consumer motives and platform use in brand-consumer interactions. *Journal of Consumer Marketing*.
- Hennig-Thurau, Thorsten, Kevin P. Gwinner, Gianfranco Walsh, and Dwayne D. Gremler (2004), "Electronic Word-of-Mouth Via Consumer Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?," *Journal of Interactive Marketing*, 18, 1, 38–52.
- Hoffman, Donna & Fodor, Marek. (2010). Can You Measure the ROI of Your Social Media Marketing?. *MIT Sloan Management Review*. 52.
- Inoue, A., Coxwell, T., & mege, M. (2020, July 20). Gymshark Marketing Strategy: How Gymshark Became the Fastest Growing Company In The U.K In 4 years. YoYoFuMedia. <https://yoyofumedia.com/blog/gymshark-marketing-strategy/>.
- Jaakonmäki, R., Müller, O., & Vom Brocke, J. (2017, January). The impact of content, context, and creator on user engagement in social media marketing. In *Proceedings of the 50th Hawaii international conference on system sciences*.
- Jansen, Jim & Zhang, Mimi & Sobel, Kate & Chowdury, Abdur. (2009). Twitter Power: Tweets as Electronic Word of Mouth. *JASIST*. 60. 2169-2188. 10.1002/asi.21149.
- Java, A., Song, X., Finin, T., & Tseng, B. (2007, August). Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 56-65).
- Kannan, Pallassana and Li, Hongshuang (Alice), Digital Marketing: A Framework, Review and Research Agenda (March 1, 2017). P.K. Kannan and Hongshuang (Alice) Li (2017) "Digital marketing: A framework, review and research agenda", *International Journal of*

Research in Marketing, 34(1), pp. 22-45., Available at SSRN:

<https://ssrn.com/abstract=3000712>

- Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business horizons*, 54(3), 241-251.
- Kim, M., & Song, D. (2018). When brand-related UGC induces effectiveness on social media: the role of content sponsorship and content type. *International Journal of Advertising*, 37(1), 105-124.
- Khan, M. L. (2017). Social media engagement: What motivates user participation and consumption on YouTube?. *Computers in human behavior*, 66, 236-247.
- Kusner, M.J., Sun, Y., Kolkin, N.I., & Weinberger, K.Q. (2015). From Word Embeddings To Document Distances. *ICML*.
- Lee, D., Kim, H. S., & Kim, J. K. (2011). The impact of online brand community type on consumer's community engagement behaviors: consumer-created vs. marketer-created online brand community in online social-networking web sites. *Cyberpsychology, behavior and social networking*, 14(1-2), 59–63. <https://doi.org/10.1089/cyber.2009.0397>
- Liu, X., Burns, A. C., & Hou, Y. (2017). An investigation of brand-related user-generated content on Twitter. *Journal of Advertising*, 46(2), 236-247.
- Long, M. (2020, November 27). *Nike, Adidas, Puma, Under Armour: Sizing up the sportswear giants in 2020*. SportsPro.
<https://www.sportspromedia.com/analysis/nike-adidas-puma-under-armour-sportswear-profit-loss-annual-reports-2020>.

- Mahdi, H. A. A., Abbas, M., Mazar, T. I., & George, S. (2015). A Comparative Analysis of Strategies and Business Models of Nike, Inc. and Adidas Group with special reference to Competitive Advantage in the context of a Dynamic and Competitive Environment. *International Journal of Business Management and Economic Research*, 6(3), 167-177.
- Ma, Liye, Baohong Sun, and Sunder Kekre (2015), "The Squeaky Wheel Gets the Grease—An Empirical Analysis of Customer Voice and Firm Intervention on Twitter," *Marketing Science*, 34 (5), 627–45.
- Malthouse, E. C., Calder, B. J., Kim, S. J., & Vandebosch, M. (2016). Evidence that user-generated content that produces engagement increases purchase behaviours. *Journal of Marketing Management*, 32(5-6), 427-444.
- Mayzlin, Dina, Yaniv Dover, and Judith Chevalier. 2014. "Promotional Reviews: An Empirical Investigation of Online Review Manipulation." *American Economic Review*, 104 (8): 2421-55.
- McGuire, William. J. (1989), "Theoretical Foundations of Campaigns," in *Public Communication Campaigns*, 2nd ed., R. E. Rice and C. K. Atkins, eds., Newbury Park, CA: Sage Publications, 43–65.
- Moon, Sangkil, Yoonseo Park, and Yong Seog Kim (2014), "The Impact of Text Product Reviews on Sales," *European Journal of Marketing*, 48 (11/ 12), 2176–97.
- Müller, J., & Christandl, F. (2019). Content is king—But who is the king of kings? The effect of content marketing, sponsored content & user-generated content on brand responses. *Computers in Human Behavior*, 96, 46-55.

- Muntinga, Daan & Moorman, Marjolein & Smit, Edith. (2011). Introducing COBRAs: Exploring motivations for Brand-Related social media use. *Contributions To Zoology - CONTRIB ZOOL.* 30. 13-46. 10.2501/IJA-30-1-013-046.
- Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), "Mine Your Own Business: Market-Structure Surveillance through Text Mining," *Marketing Science*, 31 (3), 521–43.
- O'Hern, Matthew & Kahle, Lynn. (2013). The Empowered Customer: User-Generated Content and the Future of Marketing. *Global Economics and Management Review.* 18. 22–30. 10.1016/S2340-1540(13)70004-5.
- Pentina, I., Guilloux, V., & Micu, A. C. (2018). Exploring social media engagement behaviors in the context of luxury brands. *Journal of Advertising*, 47(1), 55-69.
- Pereira, Roberto & Baranauskas, M. Cecilia & da Silva, Sergio Roberto. (2010). Social software building blocks: Revisiting the honeycomb framework. 2010 International Conference on Information Society, i-Society 2010. 253 - 258. 10.1109/i-Society16502.2010.6018707.
- Pitt, Leyland & Berthon, Pierre & Campbell, Colin. (2008). Ad Lib: When Customers Create the Ad. *California management review.* 50. 6-30. 10.2307/41166454.
- Reis, H. T., Sheldon, K. M., Gable, S. L., Roscoe, J., & Ryan, R. M. (2000). Daily well-being: The role of autonomy, competence, and relatedness. *Personality and Social Psychology Bulletin*, 26(4), 419—435.
- Rohm, A., Kaltcheva, V. D., & Milne, G. R. (2013). A mixed-method approach to examining brand-consumer interactions driven by social media. *Journal of Research in Interactive Marketing.*

- Roma, P., & Aloini, D. (2019). How does brand-related user-generated content differ across social media? Evidence reloaded. *Journal of Business Research*, 96, 322-339.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68-78
- Schivinski, B., & Dabrowski, D. (2016). The effect of social media communication on consumer perceptions of brands. *Journal of Marketing Communications*, 22(2), 189-214.
- Smith, G., *Social Software Building Blocks*.
[Http://nform.ca/publications/social-software-building-block](http://nform.ca/publications/social-software-building-block).
- Schweidel, David A., and Wendy W. Moe (2014), "Listening In on Social Media: A Joint Model of Sentiment and Venue Format Choice," *Journal of Marketing Research*, 51 (4), 387–402
- Shao, Guosong (2009). "Understanding the appeal of user-generated media: a uses and gratification perspective". *Internet Research*. **19** (1): 7–25.
 doi:10.1108/10662240910927795. ISSN 1066-2243. S2CID 17291486.
- Sheldon, K. M., Elliot, A. J., Kim, Y., & Kasser, T. (2001). What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of Personality and Social Psychology*, 80(2), 325—339.
- Shirky, C. A group is its own worst enemy: Social structure in social software. O'Reilly Emerging Technology conference. Santa Clara. CA. http://www.shirky.com/writings/group_enemy.html
- Smith, A. N., Fischer, E., & Yongjian, C. (2012). How does brand-related user-generated content differ across YouTube, Facebook, and Twitter?. *Journal of interactive marketing*, 26(2), 102-113.

Statista (2021), “Global digital population as of January 2021,”

<https://www.statista.com/statistics/617136/digital-population-worldwide/>

Statista (2021), “Most popular social networks worldwide as of January 2021, ranked by number of active users,” <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

Tirunillai, Seshadri, and Gerard J. Tellis (2014), “Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation,” *Journal of Marketing Research*, 51 (4), 463–79.

The Consumer Content Report: Influence in the Digital Age. Stackla. (2019, June 19).

<https://stackla.com/resources/reports/the-consumer-content-report-influence-in-the-digital-age/>.

Thorson, Esther, and Shelly Rodgers (2012), “What Does ‘Theories of Advertising’ Mean?,” in *Advertising Theory*, Shelly Rodgers and Esther Thorson, eds., New York: Routledge, 3–17.

van Reijmersdal, E. A., Fransen, M. L., van Noort, G., Oprea, S. J., Vandeberg, L., Reusch, S., van Lieshout, F., & Boerman, S. C. (2016). Effects of Disclosing Sponsored Content in Blogs: How the Use of Resistance Strategies Mediates Effects on Persuasion. *American Behavioral Scientist*, 60(12), 1458–1474. <https://doi.org/10.1177/0002764216660141>

Voorveld, H. A., van Noort, G., Muntinga, D. G., & Bronner, F. (2018). Engagement with social media and social media advertising: The differentiating role of platform type. *Journal of advertising*, 47(1), 38-54.

- Voorveld, Hilde A.M., Theo Araujo, Stefan F. Bernritter, Edwin Rietberg, and Rens Vliegenthart (2018), “How Advertising in Offline Media Drives Reach of and Engagement with Brands on Facebook,” *International Journal of Advertising*, 37 (5), 785–805
- Voorveld, H. A. (2019). Brand communication in social media: a research agenda. *Journal of Advertising*, 48(1), 14-26.
- Ye, Qiang & Law, Rob & Gu, Bin & Chen, Wei. (2011). The Influence of User-Generated Content on Traveler Behavior: An Empirical Investigation on the Effects of E-Word-of-Mouth to Hotel Online Bookings. *Computers in Human Behavior*. 27. 634-639. 10.1016/j.chb.2010.04.014.
- Zhu, Y. Q., & Chen, H. G. (2015). Social media and human need satisfaction: Implications for social media marketing. *Business horizons*, 58(3), 335-345.
- 3 Reasons Why User-Generated Content Works. Sprinklr. (2018, June 21).
<https://blog.sprinklr.com/three-reasons-user-generated-content-works/>