Article

Classifying and Characterizing Fandom Activities: A Focus on Superfans' Posting and Commenting Behaviors in a Digital Fandom Community

Yeoreum Lee [1] and Sangkeun Park [2],*

[1] Department of Service Design Engineering, Sungshin Women's University, Republic of Korea; yeoreum.lee@sungshin.ac.kr (Y.L.)
[2] Department of Software Convergence, Kyung Hee University, Republic of Korea; sk.park@khu.ac.kr (S.P.)
* Correspondence: Sangkeun Park (e-mail: sk.park@khu.ac.kr)

Abstract: As digital fandom communities expand and diversify, user engagement patterns increasingly shape the social and emotional fabric of online platforms. In the era of Industry 4.0, data-driven approaches are transforming how online communities understand and optimize user engagement. In this study, we examine how different forms of activity, specifically posting and commenting, characterize fandom engagement on Weverse, a global fan community platform. By applying a clustering approach to large-scale user data, we identify distinct subsets of heavy users, separating those who focus on creating posts (post-heavy users) from those who concentrate on leaving comments (comment-heavy users). A subsequent linguistic analysis using the Linguistic Inquiry and Word Count (LIWC) tool reveals that post-heavy users typically employ a structured, goal oriented style with collective pronouns and formal tones, whereas comment-heavy users exhibit more spontaneous, emotionally rich expressions enhanced by personalized fandom-specific slang and extensive emoji use. Building on these findings, we propose design implications such as pinning community driven content, offering contextual translations for fandom-specific slang, and introducing reaction matrices that address the unique needs of each group. Taken together, our results underscore the value of distinguishing multiple dimensions of engagement in digital fandoms, providing a foundation for more nuanced platform features that can enhance positive user experience, social cohesion, and sustained community growth.

Keywords: digital fandom; industry entertainment; online platforms; engagement; social interaction; superfans; user experience; service design.

Received:
Revised:
Accepted:
Published:

Citation: Lee, Y.; Park, S. Title. *Appl. Sci.* **2025**, *1*, 0. https://doi.org/

1. Introduction

Digital fandom has grown alongside the rise of online communities, fueled by technologies that enable people with shared interests to connect more easily. Fans now gather on a diverse array of digital platforms, including dedicated forums such as Reddit [1], Twitch [2], Discord [3], and Tiktok[4] as well as typical social media groups on platforms like X and Instagram [5]. Inevitably, fans' participation in these environments spans a range of media-related behaviors—views, likes, shares, posts, comments, chatting, streaming, and more—that reflect their levels of engagement.

Within these communities, users can be classified by how deeply they involve themselves in fan activities. Individuals who display exceptional commitment are often referred to as *superfans*. Superfans regularly stand out through the quantity and quality of their contributions, going beyond ordinary participation to enrich the fan ecosystem. They play pivotal roles in online interactions (e.g., discussions, idea-sharing) and offline events (e.g., community meetups), thereby strengthening connections among long-time members while also drawing in newcomers [6]. Research has extensively examined the impact of fans on community growth and sustainability [7–9], underscoring the sociocultural significance of fandom and the potential for fans to collaborate in driving entertainment industry trends.

Despite the recognized importance of superfans, most studies treat them as a homogeneous group that simply exhibits "excessive engagement" relative to casual fans, leaving a gap in understanding the nuances of superfan behavior. Existing scholarship has tended to focus on high-level metrics (e.g., hours spent, total posts made), without closely examining the varied *types* of activities or the linguistic and social characteristics that may distinguish different superfan subgroups.

This study addresses the following research questions:

1. **RQ1:** How can superfans be classified into distinct subgroups based on their activity patterns?
2. **RQ2:** How do digital fandom activities vary across different user groups?
3. **RQ3:** In what ways do linguistic styles differ among these activity-based user groups?

In order to unpack these research questions, we investigate large-scale, digitally captured fandom activities to clarify the behavioral and linguistic traits of superfans in media-rich online environments in this study. By comparing superfans with regular fans, we propose that superfans are not defined solely by the intensity of their participation but also by the distinct types of engagement they pursue with fandom activities. We focus on the BTS fandom, also known as ARMY, one of the most popular K-pop groups with particularly active digital fandom, as an illustrative case recognized for its extensive and globally connected online presence. Using a dataset of 167,456 posts and 484,437 comments from 3,410 users on Weverse (https://weverse.io/), a dedicated K-pop fan platform, we offer an empirical, data-driven view of how superfans operate within digital fandoms.

Our approach involves identifying patterns of user behavior and language use to better understand different types of superfans within digital fandom communities. This perspective allows us to provide a more detailed view of how superfans participate and communicate within these platforms. Based on these insights, we propose design strategies for HCI practitioners, social computing researchers, and platform developers who aim to foster more inclusive, engaging, and emotionally resonant fan experiences. By examining superfan activities in the context of broader community trends, this study highlights the potential of data-informed, user-centered design to enhance participation and belonging in digital fandom ecosystems.

2. Related Work

2.1. Understanding Superfans in the Context of Digital Fandoms

In the field of HCI, numerous studies have examined the socio-cultural impact of fandom, including K-pop. As fandom activities have moved online and become platform-centric, researchers increasingly draw on large-scale data analytics [19,20]. However, many foundational insights into fandom were established before the digital era, and the corresponding frameworks and classifications continue to be used today.

Although criteria for classifying fans vary, a consistent theme is that not all fans engage at the same level. Among them, *superfans* typically exhibit a relatively high level of engagement, characterized by deep community involvement, including extensive knowledge sharing, active content sharing, and various media-related activities in digital fandom. Superfans have been described as highly dedicated individuals who prioritize their fandom

Table 1. Hierarchy of Fan Engagement from Literature Reviews

#	Hierarchy	Description	Method	Reference
1	Entertainment-social, Intense-personal, Borderline-pathological	Different levels of intensity in Parasocial relationships	Survey	Maltby et al. (2006) [10]
2	Dilettante, Dedicated, Devoted	Classifying of intensity of involvement with the fandom	Survey	Thorne (2011) [11]
3	Ambient Fans, Engaged Fans, Superfan, Executive Fans	Levels of fans based on the engagement	Survey	Edlom and Karlsson (2021) [12]
4	Casual Fans, Active Fans, Hard-core Fans	Levels of fans based on the engagement	Ethnography	Fiske (1992) [13]
5	Peripheral, Marginal, Core Fans	The spectrum of fan behaviors	Ethnography, Interview	Duffett (2013) [14]
6	Passive Fans, Active Fans, Contributive Fans	Classifying of levels of participation in fandom	Interviews, Observation, Media Analysis	Abercrombie and Longhurst (1998) [15]
7	Fans, Fanatics, Extremists	Classifying of the darker side of fandom impact	Literature Review, Case Studies	Sandvoss (2005) [16]
8	Casual, Regular, Fanatic Fans	Levels of fans based on the engagement	Survey	Tapp and Clowes (2002) [17]
9	Peripheral, Core Fans	Classifying of fan engagement based on their degree of immersion	Ethnography, Qualitative Analysis	Hills and Argyle (1998) [18]

over most aspects of their daily lives. From an industry perspective, it is important to know who the superfans are who share and comment on everything because it's important to engage them in promotions [12]. Such behaviors position superfans as central figures within the fandom ecosystem, often driving discussions, spreading news, and fostering a sense of community.

A number of fan typologies have been proposed, primarily based on emotional investment and intensity of participation, as shown in Table 1. While these frameworks have proven valuable, they were developed largely through surveys, interviews, and ethnographies conducted in offline or pre-digital contexts. The evolution of fandom and the transformation of traditional social media-based platforms like Instagram and YouTube, as well as digital platforms like Weavers, TikTok, and Twitch that allow for a variety of media activities, has dramatically changed the way we interact with fans. It has also generated extensive digital traces (e.g., posts, comments, shares, and likes) that offer new insights into fandom behavior at scale—information not easily captured via traditional methods. Consequently, there is a growing need for approaches that explore digital fandom dynamics by examining fan engagement patterns. Moreover, it is crucial to investigate whether different forms of engagement, such as commenting versus posting, reflect distinct motivations and outcomes.

Several HCI and social computing studies have classified heavy users according to their activities online. For instance, some have examined question-and-answer patterns in online forums [21], or content creation and sharing in e-government contexts [22]. Others have focused on usage intensity in specific platforms, such as Instagram [23] or social chatbots [24], with the goal of improving user experiences, boosting engagement, and informing platform design. Regarding fandom, researchers have analyzed social networks [9] and the content of posts [1]. However, these studies often overlook the distinct interaction dynamics within digital fandom communities, where engagement extends beyond typical social media behaviors. Fan activities are not merely about content consumption but also involve highly participatory, emotionally charged interactions that shape community culture. This gap calls for a more granular, data-driven analysis of fan behaviors to better capture the nuances of engagement in digital fandoms.

Thus, this study adopts a data-driven approach to classify and examine fans in digital contexts by analyzing large-scale posting and commenting data. Through this method, we introduce a modern taxonomy that captures the complexities of digital fandom, emphasizing both the breadth and depth of superfan activities rather than relying solely on traditional indicators. Leveraging computational methods to study digital traces, our research contributes to a deeper understanding of fandom interaction in online environments—furthering discussions on community engagement, participation hierarchies, and user behavior in HCI.

2.2. Differences in Commenting and Posting on Social Media as Fandom Activities

Although digital fandoms operate on various platforms, core social media actions generally include viewing, liking, commenting, and posting. Among these, posting and commenting are recognized as more active forms of participation [25], and their purposes and impacts are distinct. Many studies have shown that the intent and impact of posting and commenting behaviors are different in social media use.

Posts tend to serve as the primary vehicle for conveying information or opinions, while comments function as interactive feedback mechanisms. In social media research, posting and commenting are frequently viewed as distinct behaviors in terms of visibility and user intent. Posts initiate discussions and deliver original content, whereas comments revolve around reacting to or extending the ideas presented in those posts.

HCI scholars have explored these differences in multiple contexts. Posts are often seen as original contributions—ranging from personal opinions to creative works—that reflect deeper motivations or personal contexts [26–28]. In contrast, comments usually take the form of immediate, text-based responses [26], thereby enabling higher levels of emotional expression and interactivity [29,30].

In fandom communities, these distinctions are particularly evident. Posts may showcase artwork, concert reviews, or other forms of content related to the artist [31], while comments generally convey collective enthusiasm, feedback, or emotional reactions. By providing a space for direct interaction, comments help shape community discourse, whereas posts set the stage for these discussions. Recognizing the unique roles of posting and commenting behaviors can guide platform design, foster richer fan experiences, and enhance overall digital fandom engagement.

3. Fan Community Platforms in Industry 4.0

With the advent of the Industry 4.0 era, which leverages technologies such as AI and big data to digitize and automate diverse industries, the entertainment industry (including music) has undergone a digital transformation [32]. This shift, encompassing personalized content recommendations and new forms of performance, has reshaped the nature of fandom. Rather than passively listening to music or watching performances as before, fans now actively engage with content—leaving posts and comments, watching vlogs, and participating in a variety of fandom activities. This transformation has not only diversified fandom experiences but also strengthened the sense of community [33] that existed prior to the digital era. As a result, fans have assumed a more participatory and influential role, reshaping their relationship with media producers and the entertainment industry.

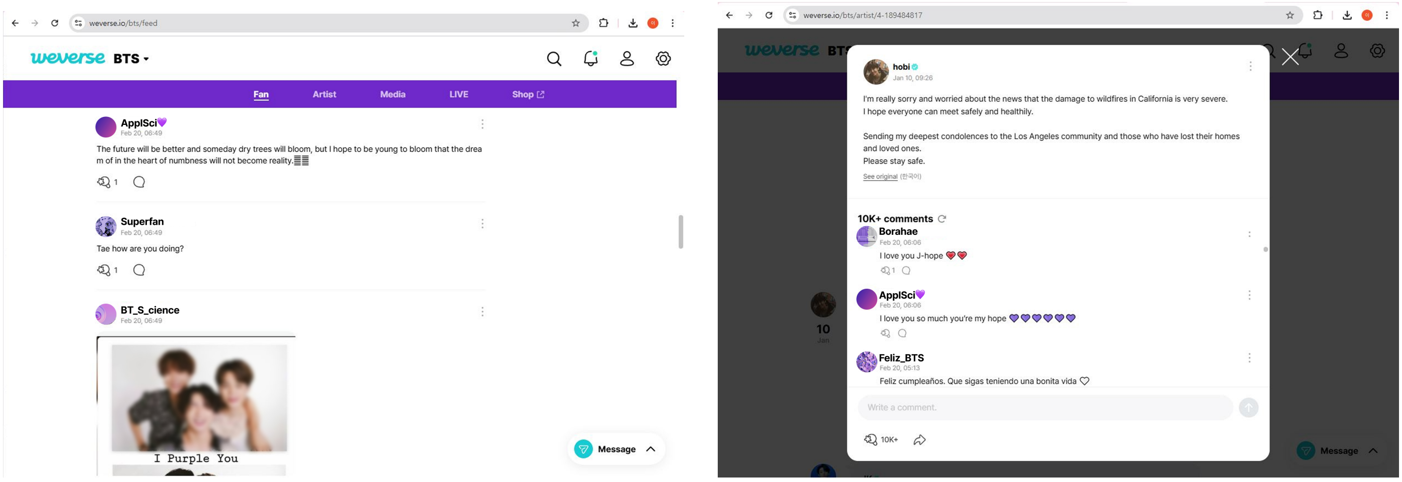
Fan community platforms aim to connect artists with fans through fan-centric content and services, typically via mobile apps. Over time, these communities have evolved in parallel with media advancements and changing fan activities [34]. In earlier stages, fan engagement was predominantly offline, as supporters gathered in person to demonstrate their dedication. With the widespread adoption of the internet, online fan forums emerged, and later, platforms such as X, YouTube, and official agency websites became key venues for fan interactions. Since 2019, as smartphone penetration in Korea reached 95%, mobile apps developed by entertainment agencies—such as Weverse and Bubble (https://www.dear-u.co/pages/business_bubble.php)—have become primary channels for artist-fan interactions. This progression highlights how fan communities adapt continuously to new technological platforms.

As media capabilities have advanced, the methods of communication in the K-pop fandom industry have diversified. Interactions between artists and fans, once hosted on online fan cafes or social media, have become even more important due to reduced offline opportunities during the COVID-19 pandemic. Various fan community platforms have evolved to include features that facilitate deeper, more direct relationships between artists and fans. A notable trend is that each organization now develops its own platform (e.g., Weverse and Bubble). Among these, Weverse has seen significant growth, reaching over 50 million registered users across 246 countries and regions by late 2022. As of early 2025, it hosts communities for 161 artists, including major K-pop groups like *BTS*, *BLACKPINK*, and *TXT*, as well as international artists such as *New Hope Club* and *Alexander 23*. This diverse lineup makes Weverse one of the platforms with the most extensive artist rosters. While direct global user base comparisons are limited, Weverse's large and diverse user engagement highlights its prominence in the global fandom platform landscape. As such, the platform is representative in terms of both the number of users and the diversity of fandoms it hosts. Focusing on *BTS*, the largest fandom on the platform, may therefore offer findings with broader generalizability. This notion is also reflected in substantial HCI research devoted to the BTS fandom, which, despite centering on a single group, has generated findings with potential generalizability to wider pop culture communities [7,8].

By 2023, Weverse reached more than 10 million MAU (Monthly Active Users) within four years of its launch (https://en.weverse.co/news/?bmode=view&idx=17441876), indicating the potential influence and rapid growth of fan community platforms. In parallel, the rise of such platforms has expanded the scope of fandom activities and introduced new social dynamics. Fans now engage more intensively in creating and sharing content, transitioning from passive consumers to active shapers of fan culture. They seek personalized experiences and direct communication with artists, supported by purpose-built features and customized content.

Examining the information structures of different K-pop platforms helps clarify the types of activities offered. Weverse functions as a comprehensive community platform, similar to Facebook, that supports a range of fan-driven actions, including posting, commenting, viewing announcements (such as official schedules), and accessing media content. We selected Weverse for this study because it supports multiple forms of fan engagement and has the largest user base among platforms developed by entertainment agencies, enabling a rich exploration of fandom behaviors.

When users open Weverse, they encounter a landing page that features a *My Communities* section, displaying the artist communities they have joined. Users can browse and select from multiple artists, and once they subscribe to a particular artist's community, it appears in this section. Weverse thus personalizes each user's experience by dynamically presenting communities that match individual preferences. The landing page serves as an entry point, allowing direct navigation to a specific artist space. Among these communities,



(a) (b)

Figure 1. (a) Screenshot Example of Comment Activity in Weverse. (b) Screenshot Example of Comment Activity in Weverse

BTS has the highest membership, with approximately 27.5 million members as of February 2025.

As illustrated in Figure 1, fans can create various types of posts through the *Fans* menu (renamed from *Feed* in October 2024) and comment on posts by other fans or by the artist under both the *Fans* and *Artist* menus. Unlike many social media platforms that rely on algorithmic feeds, Weverse displays all posts and comments strictly in reverse chronological order. This means that visibility is determined by the time of posting rather than any recommendation or moderation system. Weverse also restricts comments to text-only content, while posts can include text, images, and videos.

4. Methodology

4.1. Data Collection

Our research questions focus on understanding how superfans can be classified into subgroups and how these subgroups differ in their activities and linguistic patterns on a digital fandom platform. Therefore, we first needed to collect user posting and commenting data on Weverse. To conduct a data-driven analysis of fan activities, we crawled the Weverse BTS channel (https://weverse.io/bts/feed) on March 3, 2024, from 11:00 p.m. to 1:26 a.m. (specifically, 11:13 p.m. on March 3 to 1:26 a.m. on March 4), using Python's Selenium package (https://www.selenium.dev/). This procedure yielded 16,020 posts and 14,223 unique user IDs.

Because we aimed to investigate the behavior of established, active users rather than recent joiners, we paused data collection for a few months. On May 24, 2024, we returned to the Weverse BTS channel and accessed each previously identified user ID via its profile page at *https://weverse.com/bts/{profile_id}*. We then crawled all posts and comments these users had written over the two-month window spanning March 3 to May 3. Profiles set to *private* or belonging to *deleted* accounts were inaccessible and therefore excluded, leaving a final dataset of 3,410 users. All of these users had joined Weverse at least two months prior, remained active during that period, and maintained publicly visible profiles. We collected a total of 167,456 posts and 484,437 comments from these users. For each post or comment, the dataset includes a timestamp, text, user nickname, and user profile URL.

In the data collection process, We collected only publicly available posts and comments from Weverse, which does not provide any personally identifying information (e.g., demographic details or email addresses) to third parties. Our dataset does not include any direct identifiers that could be used to pinpoint individual users. By restricting our analysis to openly accessible content, we respect user privacy while still capturing large-scale engagement patterns.

This study was exempt from IRB review (exemption number: SSWUIRB-2024-052) because it involved collecting publicly available posts and comments without gathering personally identifiable information such as usernames or email addresses.

4.2. Data Preprocessing

Table 2. Distribution of languages in posts and comments

Language	Posts (# / %)	Comments (# / %)
English	70,921 (42.4%)	173,188 (35.8%)
Korean	30,380 (18.1%)	138,666 (28.6%)
Japanese	10,595 (6.3%)	51,837 (10.7%)
Russian	9,129 (3.3%)	18,268 (3.8%)
Tagalog	15,467 (9.2%)	13,630 (2.8%)
Arabic	994 (0.6%)	12,557 (2.6%)
Spanish	3,475 (2.1%)	2,348 (0.5%)
Other	26,495 (18.0%)	73,943 (15.3%)

Weverse is a global platform with a diverse user base, resulting in a multilingual dataset. To identify the language of each post and comment, we used *lingua-py* (https://github.com/pemistahl/lingua-py), a language-detection model optimized for short or mixed-language text. Table 2 presents the distribution of languages in the posts and comments, showing that English is most common, followed by Korean and Japanese. Note that the prominence of English or Korean text does not necessarily reflect user nationality, as many individuals use translation tools. Approximately 0.1% of posts and 5.5% of comments were labeled as *None* because they were unclassifiable (e.g., containing only emojis).

4.3. User Clustering

Table 3. Inertia and Silhouette Analysis for each k

k	Inertia	Silhouette Score
2	4256.3	0.896
3	2381.4	0.905
4	1681.0	0.870
5	1246.9	0.756
6	979.7	0.737
7	801.6	0.733

To address the research questions that require identifying distinct user groups, we draw on prior research indicating that superfans are heavy users of digital fandom platforms and deeply engaged in various platform activities. Many previous studies identify "heavy users" by applying a specific threshold (e.g., the top X%). For example, Chin et al.[35] defined the top 1% of chatbot users as superusers, while Wikipedia researchers commonly regard individuals with at least 5,000 or 10,000 edits as heavy contributors[36]. In OpenStreetMap, the 20:80 Pareto principle underpins the notion of power users, where a small subset of contributors produces the majority of edits [37].

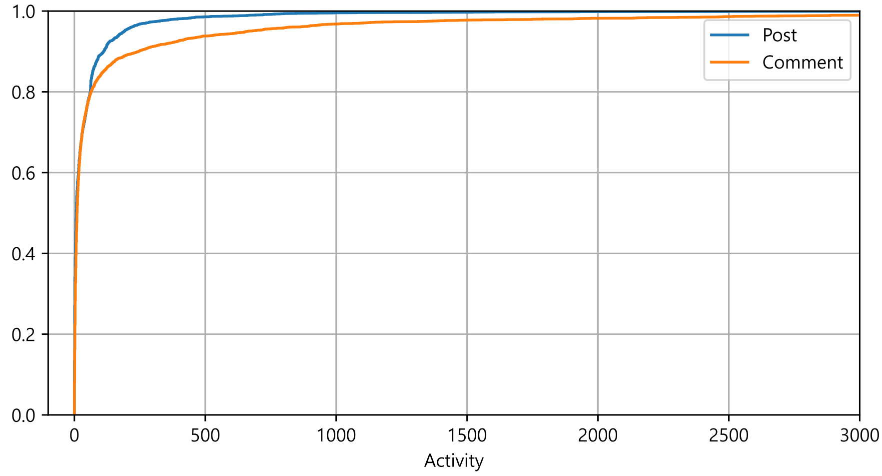


Figure 2. Cumulative Distribution Function (CDF) of the Number of Posts and Comments per User

In our dataset, however, some users predominantly posted while others focused more on commenting. A simple binary classification (e.g., general users vs. superfans) would not capture these nuanced patterns. Therefore, we applied *k*-means clustering with various *k* parameters using scikit-learn (https://scikit-learn.org/), an open-source Python machine learning library. The input features were each user's total numbers of posts and comments, both standardized through scikit-learn's *StandardScaler* to account for wide variability.

To identify the most suitable *k* value, we performed additional analyses using the elbow method and silhouette coefficients to validate our choice of the number of clusters (see Table 3). Based on the inertia values from the elbow method, we observed that the rate of decrease in inertia began to level off around *k*=3. Furthermore, our silhouette analysis showed that *k*=3 yielded the highest silhouette coefficient (0.905), indicating well-defined cluster separation. As a result of applying *k*-means clustering with *k*=3, 3,301 users (96.8%) were classified as *general*, 71 (2.1%) as *comment-heavy*, and 38 (1.1%) as *post-heavy*, thereby revealing two distinct types of superfans.

4.4. Linguistic Analysis

To explore how linguistic styles vary across activity-based user groups, we conducted a linguistic analysis using the Linguistic Inquiry and Word Count (LIWC), specifically the LIWC-22 version (https://www.liwc.app/), a validated psycholinguistic dictionary [38,39]. LIWC assesses stylistic, tonal, and content-centric features of fan-generated text by categorizing words into psychologically meaningful groups [40], such as emotional expression, cognitive processes, and social interaction cues. Widely used in social computing research to analyze social media posts [41,42] and online conversations [35], LIWC provides a systematic and quantitative method for examining language use. In this study, we applied LIWC-22 to both posts and comments to better understand how fans communicate and express themselves on the Weverse platform.

5. Result

In this section, we present the results of our analysis on users' posting and commenting behaviors and examine how these behaviors relate to linguistic usage. We also provide a deeper exploration of how these findings vary across the three user categories: the *general* group, the *comment-heavy* group, and the *post-heavy* group.

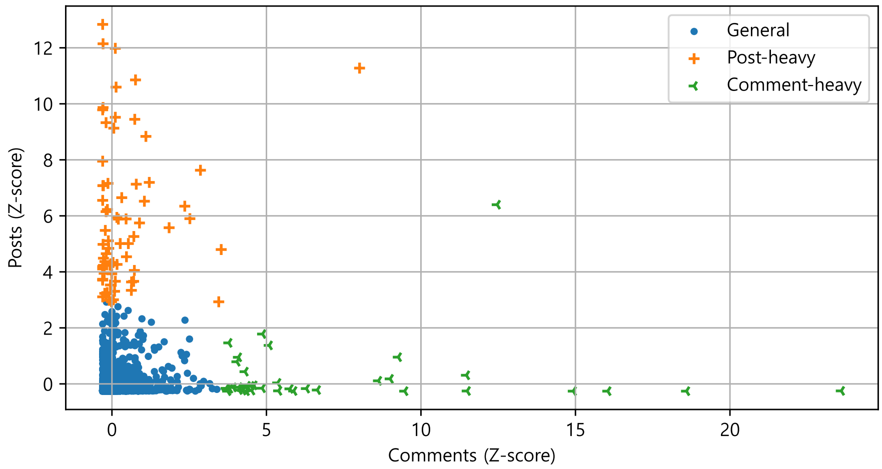


Figure 3. A result of *k*-means Clustering of Users based on Posting and Commenting Activities

Table 4. The Average Post and Comment Per User Across Groups

	General	Post-heavy	Comment-heavy
Post	M=34.2 (SD=65.9)	M=1209.7 (SD=767.9)	M=119.1 (SD=205.9)
Comment	M=71.3 (SD=189.8)	M=281.8 (SD=633.9)	M=3355.7 (SD=1,410.0)

5.1. User Behavioral Patterns

We found that the distribution of users' posts and comments follows a long-tail (Pareto) distribution, as shown in Figure 2. Although most users contribute only a small number of posts or comments, a relatively small group of heavy users makes a substantial contribution. Specifically, for posts, the top 20% of users account for 137,678 posts (82.2% of all posts), and for comments, the top 20% of users account for 450,711 comments (93.0% of all comments). The maximum number of posts is 3,913, and the maximum number of comments is 7,174.

5.1.1. User Clustering: Identifying General Users and Two Distinct Superfan Types

To address RQ1, ("How can superfans be classified into distinct subgroups based on their activity patterns?"), we conducted user clustering and identified 3,301 users (96.8%) as general, 71 (2.1%) as comment-heavy, and 38 (1.1%) as post-heavy, revealing two distinct types of superfans. Figure 3 shows the results of clustering users according to their posting and commenting activities. Because the number of posts and comments varies greatly among users, we applied scikit-learn's StandardScaler to transform the data so that each feature has a mean of 0 and a standard deviation of 1 before performing the clustering. As a result, the ticks on the X-axis and Y-axis are displayed as Z-scores for posts and comments, respectively. We observed that post-heavy and comment-heavy users diverged from the general user cluster around a Z-score of 3, which is a commonly recognized outlier threshold [43]. These heavy groups exhibit distinct behavioral tendencies. For example, most post-heavy users concentrate on posting, while most comment-heavy users focus on commenting. This indicates that heavy users are typically active in one type of activity rather than both, as also shown in Table 4.

5.1.2. Posting and Commenting Behaviors across General Users and Superfan Groups

To address RQ2 ("How do digital fandom activities vary across different user groups?"), we examined the posts and comments and discovered numerous instances

Table 5. Overall Data

Clustering	Users	Posts		Comments	
		Count	Users	Count	Users
General	3301	113,034	2,858	235,477	3,032
Comment-heavy	71	8,455	62	238,253	71
Post-heavy	38	45,967	38	10,707	32
Total	3,410	167,456	2,958	484,437	3,135

Table 6. Deduplicated Data

Clustering	Users	Posts		Comments	
		Count	Users	Count	Users
General	3,301	92,087	2,858	219,989	3,032
Comment-heavy	71	6,191	62	212,356	71
Post-heavy	38	20,114	38	10,455	32
Total	3,410	118,392	2,958	442,800	3,135

of exact text duplication. When we remove duplicate text from the same user, the total number of posts and comments decreases substantially, as shown in Table 5 and Table 6. However, this effect is more pronounced for posts than for comments. For example, among post-heavy users, the original 45,967 posts drop to 20,114 after deduplication, a reduction of more than 50%. This decrease is much larger than that observed for general users or comment-heavy users, indicating that post-heavy users frequently submit identical posts multiple times. Many of these are large-scale promotional posts, referred to as a mass fan campaign, in which fans coordinate a particular activity such as information sharing, vote promotion, or stage whisper. For instance, some post-heavy users posted numerous invitations from third-party groups discussing BTS, as follows:

"We would like to invite Y'll to BPB SCOUTS on TELEGRAM. JOIN: https://... let's have fun talking about our BANGTAN BOY'S and many more stuffs. This groups is safe and secure." (posted more than 1,000 times by several post-heavy users)

"Hello dear ARMY we created the group TELEGRAM for BTS ARMY GIRLS. To enjoy chatting with many ARMYs including foreign countries. We are there from all over the world. We are talking and sharing all the information, links, voting issues related to BTS". (posted more than 1,000 times by several post-heavy users)

Another example involves a promotional campaign that encourages fans to vote for a newly released BTS song or participate in a popularity contest, as follows:

"please Vote for V - FRI(END)S song. please Vote for V - FRI(END)S song. Make him wins. we can do it fighting. https://en.fannstar.tf.co.kr/rank/view/wmusic ..." (posted 26 times by one post-heavy user)

"!!!ARMY!!! please vote for TAEHYUNG - the most handsome man in the world 2024. He's now second position..." (posted 182 times by one post-heavy user)

We also found that many posts were nearly identical, differing by only a few characters. For example, some post-heavy users repeatedly posted countdowns to a BTS member's military discharge or tracked BTS's daily Spotify streaming numbers as a form of stage

Table 7. English-text Filtered Data

Clustering	Users	Posts		Comments	
		Count	Users	Count	Users
General	2,727	42,325	2,120	100,400	2,270
Comment-heavy	71	3,070	53	96,814	71
Post-heavy	37	14,515	36	7,551	32
Total	2,835	59,910	2,209	204,765	2,373

whisper, as follows:

*"I WAITING FOR YOU. MY LOVE "D-Day" on Spotify Counter Day **326** — 2,496,831 ..."*

In contrast, removing duplicate text from comments does not result in as large a reduction as observed for posts. Even for post-heavy users, the total number of comments remains relatively stable after deduplication. Although general and comment-heavy users show about a 10% decrease, post-heavy users exhibit only minimal reduction. This finding indicates that post-heavy users frequently re-post identical content but rarely duplicate their comments. We then investigated the duplicated comments and found that they tend to be simple, common responses. For example:

"what are you doing now" (commented twice by a comment-heavy user)

"happy birthday" (commented twice by a comment-heavy user)

Overall, repeated content is more prevalent in posts than in comments, likely because the feed updates extremely quickly. As a result, users often re-post identical or very similar content to maintain visibility in the rapidly changing feed.

5.2. Linguistic Usage Patterns

To address RQ3 ("In what ways do linguistic styles differ among these activity-based user groups?"), we used the Linguistic Inquiry and Word Count (LIWC) tool to examine the stylistic and tonal features of user-generated text. For practical and methodological reasons, we filtered and deduplicated only the English-language posts and comments (see Table 7). Although Weverse is multilingual, English is the most commonly used language on the platform (42.4% of posts and 35.8% of comments; see Table 2), functioning as a global lingua franca that can be accessed and understood by a broad international audience. We also observed that English-language posts and comments tend to have the largest influence in the community, since non-English messages can be less accessible to users who do not speak those languages.

Focusing on English-language content avoids the additional complexity of translations, which may introduce unintended shifts in nuance or style. Meanwhile, carrying out linguistic analyses in each non-English language would require dictionaries or models attuned to different cultural and grammatical structures, potentially leading to incomparable results. Given these considerations, we decided to conduct our analyses on the widely used and globally accessible English texts. In total, this process yielded 59,910 posts from 2,209 users and 204,765 comments from 2,373 users; of these, 2,722 were general users, 71 were comment-heavy, and 37 were post-heavy. We used only these English posts and comments for the linguistic analysis.

We then applied LIWC-22 to these English-language posts and comments. Some posts and comments could not be analyzed by LIWC because they contained no text or were too

Table 8. A result of LIWC analysis (1. *Post* vs. *Comment*, 2. *Post* comparision across groups, and 3. *Comment* comparision across groups)

LIWC	All Users (N=2,747)			Users who **POST**ed at least once					Users who **COMMENT**ed at least once				
	Post	Comment	*p* *(t-test)*	General (N=2,029)	Comment -heavy (N=49)	Post -heavy (N=34)	*p* *(ANOVA)*	*p* *(post-hoc)*	General (N=2,193)	Comment -heavy (N=71)	Post -heavy (N=32)	*p* *(ANOVA)*	*p* *(post-hoc)*
Count	55,078	195,118		38,936	2,756	13,386			95,668	92,038	7,412		
Analytic	**33.78**	**96.2%**	***	**32.65**	**89.3%**	**115.8%**	***	***	**33.01**	**94.9%**	**122.6%**	***	***
Clout	**66.45**	**106.5%**	***	**68.51**	**107.9%**	**86.3%**	***	***	**67.90**	**106.6%**	**124.4%**	***	***
Authentic	**60.09**	**79.8%**	***	**61.90**	**97.1%**	**88.9%**	***	***	**48.27**	**102.8%**	**49.3%**	***	***
Tone	**87.45**	**100.5%**	***	**87.43**	**102.2%**	**99.6%**	**	**	**86.07**	**104.4%**	**98.3%**	***	***
Personal pronoun	**14.48**	**93.3%**	***	**14.91**	**116.3%**	**84.8%**	***	***	**13.02**	**108.0%**	**98.8%**	***	***
I	6.82	86.6%	***	6.74	128.0%	99.2%	***	***	5.75	109.8%	50.3%	***	***
We	1.02	106.8%	*	1.01	127.1%	100.3%	***	**	0.96	95.3%	519.1%	***	***
You	5.54	104.7%	***	6.10	105.2%	61.2%	***	***	5.54	111.6%	80.8%	***	***
She/He	0.68	64.3%	***	0.63	88.7%	131.7%	***	***	0.47	87.5%	51.5%	***	***
They	0.29	61.0%	***	0.30	75.2%	97.2%			0.20	74.6%	103.8%	***	***
Drives	**4.68**	**117.0%**	***	**4.45**	**164.9%**	**108.2%**	***	***	**5.14**	**105.3%**	**207.8%**	***	***
Affiliation	2.40	138.2%	***	2.25	150.8%	115.3%	***	***	2.86	120.6%	261.2%	***	***
Achieve	0.71	68.3%	***	0.67	91.2%	125.0%	***	***	0.45	102.8%	241.1%	***	***
Power	1.59	106.9%	***	1.53	219.6%	92.4%	***	***	1.85	81.7%	114.8%	***	***
Affect	**12.87**	**92.3%**	***	**13.56**	**96.3%**	**79.8%**	***	***	**10.71**	**125.7%**	**69.0%**	***	***
Positive Tone	11.89	92.4%	***	12.58	95.0%	78.4%	***	***	9.75	129.4%	69.4%	***	***
Negative Tone	0.79	93.7%	***	0.79	108.2%	93.6%	***	***	0.80	85.0%	66.9%	***	***
Emotion	6.28	69.2%	***	6.57	86.3%	84.7%	***		4.04	119.9%	53.0%	***	***
Swear	0.03	74.7%	***	0.04	215.1%	44.3%	**	**	0.03	93.8%	41.2%	*	
Social	**20.78**	**100.2%**	***	**21.55**	**122.0%**	**80.7%**	***	***	**19.45**	**113.2%**	**121.9%**	***	***
Social Behavior	7.86	102.4%	***	7.91	104.8%	96.3%	*	*	7.35	119.0%	111.8%	***	***
Social References	11.48	101.6%	***	11.75	133.7%	83.6%	***	***	11.07	108.4%	137.5%	***	***
Conversation	**2.05**	**151.7%**	***	**2.04**	**97.2%**	**102.2%**			**3.48**	**81.7%**	**51.4%**	***	***
Netspeak	1.61	133.1%	***	1.68	96.2%	84.9%	***		2.47	76.2%	49.5%	***	***
Assent	0.30	274.3%	***	0.21	61.6%	291.7%	***	***	0.81	107.9%	54.5%	***	***
Non-fluencies	0.18	154.2%	***	0.17	113.2%	108.3%			0.32	76.9%	53.1%	***	***
Filter	0.12	125.5%	***	0.14	126.7%	54.1%	***	**	0.18	72.5%	54.6%	***	
Emoji	**51.48**	**92.2%**	***	**51.15**	**126.5%**	**97.2%**	***	***	**44.01**	**120.2%**	**54.6%**	***	***

* Green cells represent higher ratios compared to the reference group or baseline (i.e., an increase in the given linguistic feature), whereas red cells represent lower ratios (i.e., a decrease in the given linguistic feature).

short for textual analysis. For instance, some posts included only a hashtag (for example, #BTS) along with images or videos, while some comments consisted solely of emojis (for example, 💜, which is a symbol of unity and loyalty for BTS fans). We, therefore, removed any posts and comments that were not analyzable by LIWC from the English-filtered dataset.

In total, we analyzed 55,078 posts and 195,118 comments submitted by 2,747 users (see Table 8). For comparisons between posts and comments, the value for comments is presented as a ratio (%) of the LIWC value for posts, and the *p (t-test)* column indicates the significance level from the t-test.

For comparisons among user groups, the LIWC values for comment-heavy and post-heavy users are presented as ratios (%) relative to those for general users. The *p (ANOVA)* column indicates the significance level obtained from the ANOVA of differences among the three groups, and the *p (post-hoc)* column shows the significance level from Tukey's HSD specifically comparing the comment-heavy and post-heavy groups. All significance levels are based on a 95% confidence level ($\alpha = 0.05$).

LIWC produces four summary variables, 1) **Analytic** [44–46], 2) **Clout** [47–49], 3) **Authentic** [50,51], and 4) **Tone** [52,53]. These values are calculated based on the presence of sufficient textual information and are standardized scores converted to percentiles (from 1 to 99). They are considered "nontransparent" dimensions in the LIWC output.

- **Analytic**: Measures the degree to which a text suggests formal, logical, and hierarchical thinking. Higher scores correlate with academic-style reasoning and organized expression. Lower scores suggest a more personal, friendly tone.
- **Clout**: Reflects the speaker's relative social status, confidence, or leadership style. Higher scores often indicate a more confident, authoritative tone with fewer personal pronouns, whereas lower scores reflect a humble or tentative communication style that includes more self-references.
- **Authentic**: Assesses how open, honest, and personal a text appears to be. High scores reflect spontaneous, unfiltered communication, while low scores often appear in prepared remarks or socially cautious statements.
- **Tone**: Measures emotional tone by combining positive and negative emotional indicators into a single dimension. Higher scores indicate more positive language, and scores below 50 suggest a more negative emotional tone.

We also considered additional LIWC dimensions that calculate the percentage of total words belonging to specific dictionaries for each dimension:

- **Personal pronoun**: Tracks the usage of personal pronouns (e.g., "I," "we," "you," "she/he," and "they").
- **Drives**: Measures expressions of motivation or goals such as affiliation, achievement, and power. This dimension includes words associated with *affiliation* (e.g., "we," "our," "us"), *achievement* (e.g., "work," "better," "best"), and *power* (e.g., "own," "order," "allow").
- **Affect**: Captures words related to emotional states, including *Positive and negative tones*, *Emotion* (e.g., "happy," "joy," "sad," "angry") and *Swear* (e.g., "shit," "damn").
- **Social**: Encompasses social processes and words associated with *prosocial behaviors* (e.g., "thanks," "love," "care") and *social references* (e.g., "parent," "friend," "his/her").
- **Conversation**: Captures words that reflect direct, interactive discourse typical of natural conversations, including *netspeak* (e.g., "lol," "haha"), *assent* (e.g., "yeah," "ok"), *non-fluencies* (e.g., "oh," "um"), and *filler* (e.g., "you know," "wow").
- **Emoji**: Counts the number of emojis relative to the total number of words in the text.

Based on these ten primary LIWC dimensions, we compared linguistic features in posts versus comments and across user groups.

5.2.1. Comparison of Linguistic Patterns in Posts and Comments

Before examining how linguistic styles vary across different user groups, we first explored the linguistic differences between posts and comments. A t-test revealed significant differences between posts and comments for all LIWC primary dimensions and sub-dimensions ($p < 0.05$ to $p < 0.001$). However, this is likely due to the large sample size. Instead of focusing on all statistically significant differences, we highlight the dimensions with larger effect sizes. In general, the substantial dimensional differences between posts and comments were found on *Authentic*, *Drives*, and *Conversation* of the primary dimensions.

Authentic showed the substantial difference, with an average score of 58.98 for posts versus 47.04 (79.8%) for comments. the Authentic dimension reflects how people reveal themselves in what they say and how they say it. Texts that score high on Authenticity are often spontaneous and unfiltered, whereas texts that score low tend to be more prepared or socially cautious. For example, public-facing posts sometimes demonstrate higher spontaneity than comments. This may occur because posts are accessible to a broad, unspecified audience, while comments target a more specific listener or topic, leading users to be more cautious in their remarks.

Drives showed a substantial difference, with an average score of 4.68 for posts versus 5.48 (117.0%) for comments. The Drives dimension encompasses language related to

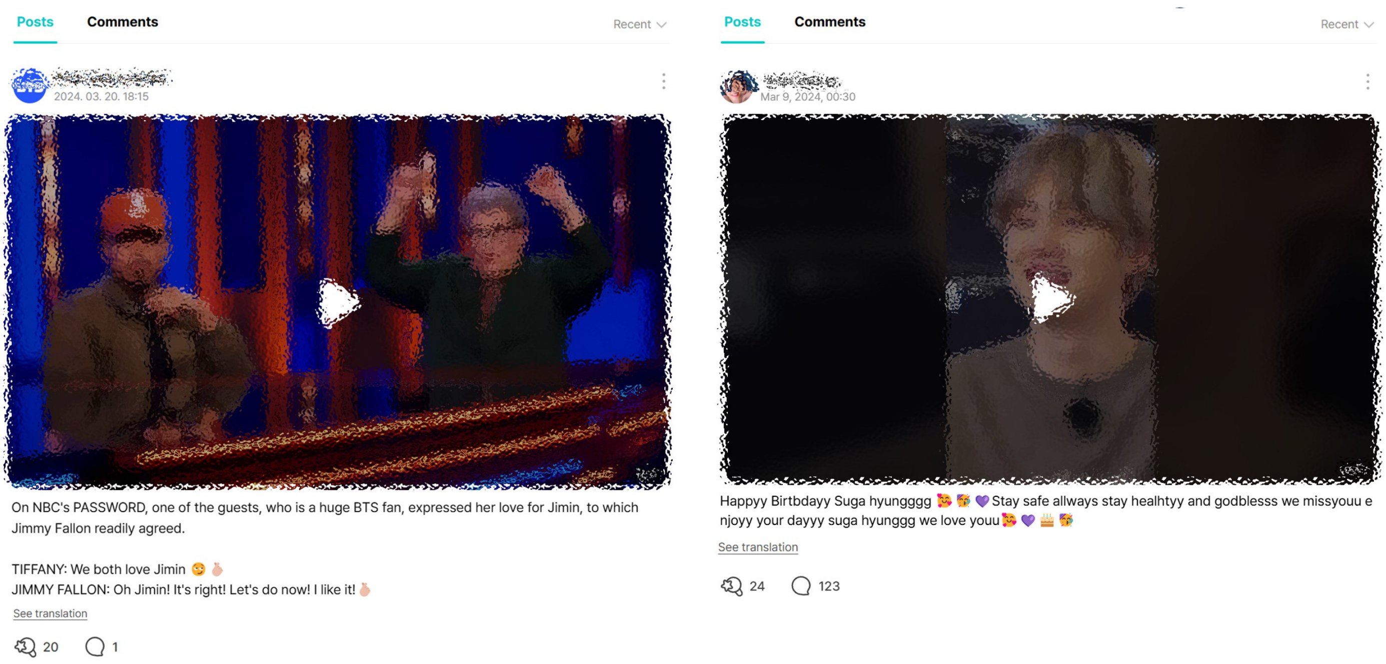


Figure 4. Examples of posts from a post-heavy user (left) and a comment-heavy user (right)

personal and social motivation, such as affiliation, achievement, and power. Higher Drives scores often indicate goal-oriented communication, where speakers explicitly express their needs or desires to influence others. In the context of comments, this might manifest as calls to action (e.g., "Let's vote," "We need to vote together"), demonstrations of pride in accomplishments, or encouragement of collective goals. Because comments often involve direct interaction with other users, it is perhaps natural that people would use more motivational or goal-focused language in this setting.

Conversation showed the greatest difference among the primary dimensions, at 2.05 for posts versus 3.11 (151.7%) for comments. The Conversation dimension captures words and expressions characteristic of direct, interactive discourse, including netspeak, assent, non-fluencies, and filler words. Elevated scores suggest a more immediate, dialogic style resembling real-time interaction. On social media, comments often function as back-and-forth exchanges, where users spontaneously respond to others' posts or comments. As a result, comments included more conversational markers, playful interjections, and informal language than posts.

5.2.2. Linguistic Differences Across User Groups

We also investigated linguistic differences among the three user groups (general, comment-heavy, and post-heavy). We conducted a one-way ANOVA for each dimension, followed by post-hoc Tukey's HSD tests whenever the ANOVA indicated significant differences. To emphasize the differences in heavy groups, we reported only the corresponding p-values in the *p (post-hoc)* column in Table 8. The results show that these heavy groups differ significantly in most dimensions. Through these findings, we confirmed that there are substantial differences in language usage patterns between the post-heavy group and the comment-heavy group within both posting and commenting contexts. Representative examples of posts and comments from post-heavy and comment-heavy users are shown in Figure 4 and Figure 5, respectively.

Analytic: Post-heavy users tended to produce more structured and formal text, indicated by higher Analytic scores in LIWC. Their writing often included well-organized sentences and collective pronouns (for example, "we" or "let's"), suggesting official an-

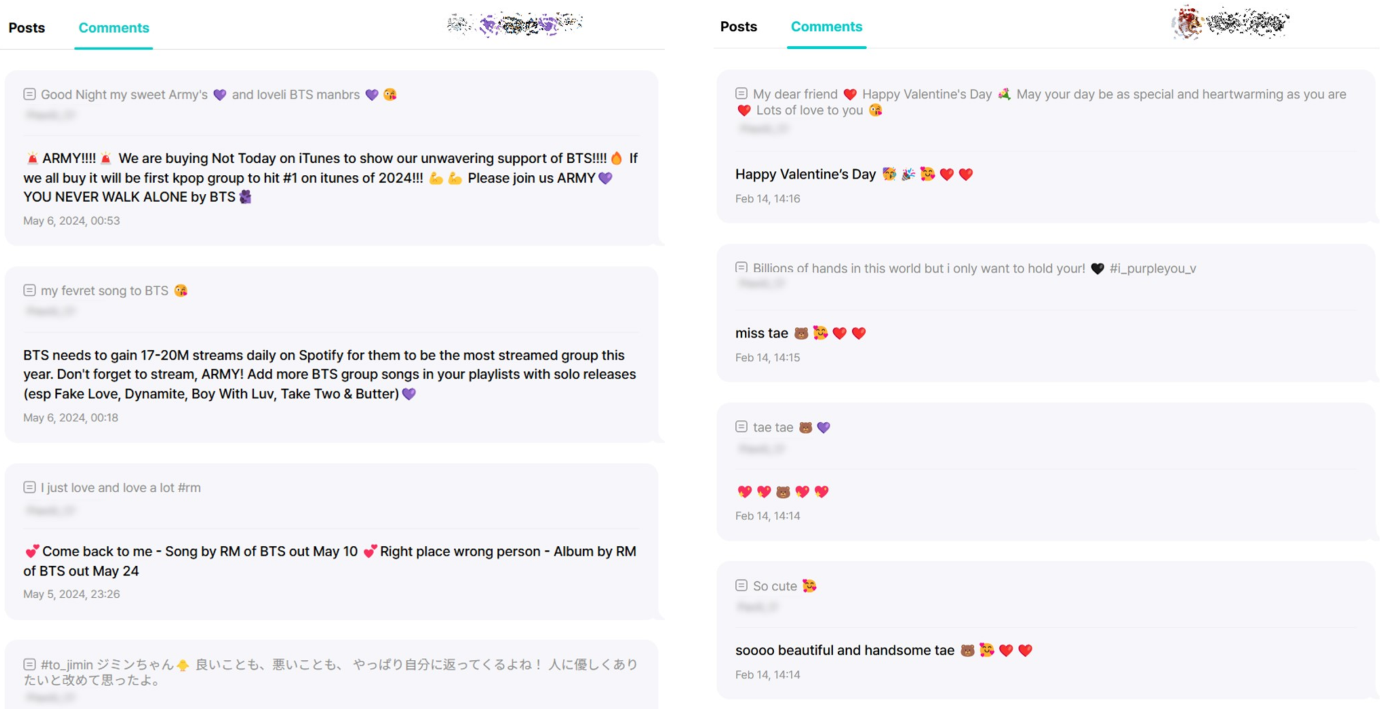


Figure 5. Examples of comments from a post-heavy user (left) and a comment-heavy user (right)

nouncements or group-focused statements. In contrast, comment-heavy users exhibited a more casual style, occasionally ignoring conventional grammar (for example, "I loooooove you") and engaging in rapid, conversational exchanges.

Clout: LIWC defines Clout as an index of social dominance or leadership in communication. Comment-heavy users displayed higher Clout when writing public posts, often by initiating conversations or making bold statements (for example, "I love you!"). In contrast, post-heavy users generally showed higher Clout in comment sections, frequently directing discussions with calls to action (for example, "Why don't we. . . ," "Let's vote. . . ").

Authentic: Although posts in general appeared more Authentic than comments across the entire user base, post-heavy users' own posts were somewhat lower in Authentic scores. This may occur because many of their posts resembled official announcements rather than spontaneous personal expressions. On the other hand, comment-heavy users' posts included more informal, candid language, which correlated with higher Authentic scores.

Tone: Comment-heavy users showed a significantly higher Tone score than general and post-heavy users, although the difference between comment-heavy and post-heavy users was not significant. This likely occurred because comment-heavy users frequently employed more positive or affectionate language (for example, "miss you," "love you").

Personal pronoun: Analysis of personal pronoun usage showed that comment-heavy users used more instances of "you," whereas post-heavy users used more instances of "we," especially in their posts. This finding suggests a more interpersonal communication style among comment-heavy users, while post-heavy users tend to emphasize collective goals or a group identity.

Drives: Post-heavy users exhibited higher Drives scores, suggesting a stronger focus on goals, leadership, or group motivation. Their posts often contained group-oriented language ("we," "let's") or directives. In contrast, comment-heavy users demonstrated a more informal communication style, characterized by frequent use of "you," more affective expressions, and a greater inclination to express empathy or affection.

Affect: Further distinctions emerged in Affect, where post-heavy users demonstrated markedly lower emotional expression overall, particularly in their comments. Compared to the other groups, which used more varied emotional terms or friendly gestures (for example, “love you,” emojis), post-heavy users maintained a more reserved tone in both posts and comments.

Social: Post-heavy users showed significantly lower Social scores than comment-heavy users when posting. This occurred because post-heavy users often focused on official announcements, such as voting promotions and information-sharing, whereas comment-heavy users spent more time interacting with one another in the post feed. However, for comments, the gap between post-heavy and comment-heavy users was smaller, though still statistically significant.

Conversation: Comment-heavy users exhibited significantly higher Conversation scores than post-heavy users, indicating more frequent use of informal conversational markers (e.g., Netspeak, assent, filler words). This pattern suggests a greater tendency toward quick, interactive exchanges. Conversely, post-heavy users displayed a more planned, topic-driven, or formal posting style reminiscent of short articles or announcements.

Emoji Usage: Across all user groups, emoji usage was high relative to total word counts in both posts and comments. When comparing across groups, comment-heavy users employed significantly more emojis in their posts than the other two groups, potentially reflecting their preference for the informal and expressive communication style commonly seen in commenting. In contrast, no significant difference was found between general users and post-heavy users in their emoji usage in posts. However, in comments, post-heavy users used emojis at roughly half the rate of general and comment-heavy users, indicating a more restrained emotional tone in that context.

To further explore this pronounced use of emojis, we additionally analyzed the types of emojis used. Interestingly, the most frequently used emoji by far was the purple heart emoji (💜), not to be confused with the standard red heart, appearing in 18,595 posts (31.0%) and 97,838 comments (47.5%). In posts, the second and third most used emojis were the ’sob’ emoji (😭) with 3,489 instances, and the ’red heart’ (❤️) with 6,729 instances. In comments, the ’red heart’ emoji (❤️) ranked second (27,975 comments), followed by the ’smiling face with 3 hearts’ emoji (🥰) with 20,820 comments.

6. Discussion

6.1. Identifying Superfans: Classifying Heavy Users by Activity Types

While traditional fan classifications from previous fandom studies have laid a valuable foundation, they are increasingly inadequate for understanding the complexities of digital fandom in the age of big data and real-time engagement. Today, fan activities leave behind a wealth of digital traces, offering more nuanced and multi-dimensional insights into fan behavior. To reflect this shift, a modern framework must go beyond one-dimensional thresholds and incorporate quantitative data that considers multi-activity engagement—particularly the content creation and interaction patterns that shape digital fandoms. Consequently, this study aims to develop a data-driven classification approach that better captures the reality of fandom in the digital age.

In this study, we applied a k-means clustering approach to identify user groups and found that superfans on digital fandom platforms can be categorized into two distinct types of heavy users based on posting and commenting behaviors. We also observed that a small subset of heavy users contributes a disproportionately large share of overall activity, aligning with the power-law distributions observed in previous research on Wikipedia and other platforms [21,35,54,55]. Traditional approaches to identifying heavy users have generally relied on threshold-based or percentile-based metrics, such as labeling the top 1% or 5% of users as superusers. For instance, Wang et al. [55] defined the top 5% of Quora users by follower count as “super users,” and later compared their voting and content contributions with those of general users. While effective, these methods often treat activity as a single dimension (for example, the number of edits or total votes).

In our study, we moved beyond this by adopting a clustering-based method (k-means) to capture multiple dimensions of activity. Specifically, we distinguished between posting and commenting behaviors instead of relying on an arbitrary cutoff. This approach also addressed the complexities of multi-criteria thresholds, which would otherwise require separate cutoffs for each type of activity. By using two primary activity types, posts and comments, we successfully identified three distinct groups: general users, post-heavy users, and comment-heavy users. These clusters exhibited significant differences in both behavioral and linguistic characteristics.

Our results suggest that treating posting and commenting simply as a single "engagement score" could mask meaningful subgroups. Instead, clustering illuminated how users predominantly gravitate toward one form of participation or another. This finding underscores the importance of considering the platform's structural features and user interaction patterns. It also complements prior work (e.g., Aldous et al. [25]), where different engagement tiers were identified based on activity level, but without explicitly separating posting and commenting.

6.2. Post-heavy Users: Official Campaigns and Structured Communication

In Section 5.1, our posting and commenting behavior analysis revealed that post-heavy users frequently share large volumes of nearly identical posts, resulting in more than a 50% reduction in the total post count after deduplication. This behavior is reminiscent of "mass fan campaigns," in which fans repeatedly post promotional content—such as voting reminders or calls for collective action—to maintain visibility in a rapidly updating feed. On social media platforms like Twitter, however, duplicated or repeated posting is generally viewed as undesirable behavior, often classified as spam or overt promotion rather than standard user activity [56,57]. Moreover, such duplicated content risks being flagged by platform algorithms as originating from false or fake accounts [58,59].

Despite their repetitive nature, the posts submitted by post-heavy users in fan communities generally serve as announcements or calls to action (for example, encouraging fellow fans to vote), rather than mere spam, as illustrated in our linguistic analysis (see Section 5.2).

A closer examination of linguistic patterns reveals that post-heavy users often employ a highly analytic and structured style, suggesting a more formal and carefully organized approach to fandom engagement. This aligns with previous findings highlighting analytic language use as indicative of a deliberate, cognitively engaged style, often associated with professional or authoritative communication [47]. Additionally, the frequent use of collective pronouns, particularly '*we*', resonates with established theories on collective identity within fandom communities, underscoring a strong sense of shared identity and common goals [60]. Moreover, lower scores observed in the Authentic dimension among post-heavy users suggest a strategic self-presentation similar to online influencers or public figures who carefully manage their persona to align with audience expectations and community norms [47,61]. Correspondingly, minimal use of affective expressions and emojis points to emotional restraint, reflecting their prioritization of clear, direct communication aimed at efficiently disseminating important fandom-related information rather than informal emotional exchanges.

Conversely, when examining their commenting behaviors, post-heavy users displayed notably higher Clout and Drives scores coupled with elevated scores in the Social dimension. Theoretically, higher Clout indicates a perceived authority and confidence, which in this context signifies that their comments often serve as guiding or coordinating contributions within community discussions [62]. The increased Drives scores, reflecting motivations and goal-driven narratives, emphasize their instrumental engagement in community management and fan-driven campaigns (for example, “Let’s vote…” and “Why don’t we…”), aligning closely with strategic user behaviors documented in digital fandom research [63]. Additionally, the increased Social dimension scores in comments highlight their active participation in social interactions and their adaptability to less formal, yet strategically important conversational contexts. Despite maintaining relatively lower Authentic and Affect scores across posts and comments, their linguistic adaptability suggests sophisticated social and communicative strategies that dynamically adjust based on interaction contexts, mirroring multifaceted communication patterns typical within online fandom communities [64]. This nuanced linguistic adaptability provides deeper insight into the strategic roles that superfans assume, extending beyond mere engagement metrics to illuminate underlying user intentions and community dynamics.

6.3. Comment-heavy Users: Social and Emotional Engagement

On the other hand, our linguistic analysis reveals that comment-heavy users typically adopt a more casual, socially oriented engagement style (see Section 5.2). Their comments are characterized by informal structures, colloquial language, and playful spellings such as “loooooove,” aligning with lower Analytic scores. Additionally, they frequently use personal pronouns (“I,” “you”) and an abundance of emojis. This informal and expressive tone contributes significantly to their higher Tone scores—indicating predominantly positive emotional language—and the frequent expression of affection, exemplified by phrases such as “I miss you” or “I love you.”

Consistent with these observations, recent large-scale studies have found that user comments across various platforms have become linguistically simpler and shorter over time, yet continually incorporate novel linguistic elements such as new slang and memes [65]. Emojis, in particular, function as a form of “social media paralanguage,” enabling commenters to express nuanced emotional tones and foster affiliative bonds, thereby strengthening community identification [66]. These findings align closely with our data, demonstrating how comment-heavy users utilize expressive markers like emojis, informal spellings, and personal pronouns to sustain a shared identity and sense of community belonging. Moreover, prior research suggests that emotional expressiveness itself acts as a catalyst for community engagement [67], highlighting how the emotionally charged and vibrant comments from this subgroup can encourage ongoing participation and strengthen communal ties.

Notably, when comment-heavy users transition to posting, their Clout and Drives scores increase relative to general users. Higher Clout scores indicate that comment-heavy users confidently position themselves, leveraging the trust and rapport established through frequent informal interactions to influence or mobilize community actions [62]. This strategically positioned influence aligns with the documented roles of fandom leaders or influential participants who shape community activities, narratives, and collective sentiments through regular social interactions [60,63]. Elevated Drives scores further suggest their posts reflect strong personal motivations aimed at mobilizing community engagement, reinforcing earlier findings that active commenters significantly influence group dynamics and sentiment cohesion within fandom communities [60,64].

Additionally, their posts exhibit higher Social dimension scores compared to post-heavy users, consistent with their inherently interactive and relationship-oriented communication style. This relational focus—even in original posts—highlights their essential role in fostering community cohesion and interpersonal connectivity within digital fandoms. Such roles are consistent with previous research identifying highly interactive fans as criti-

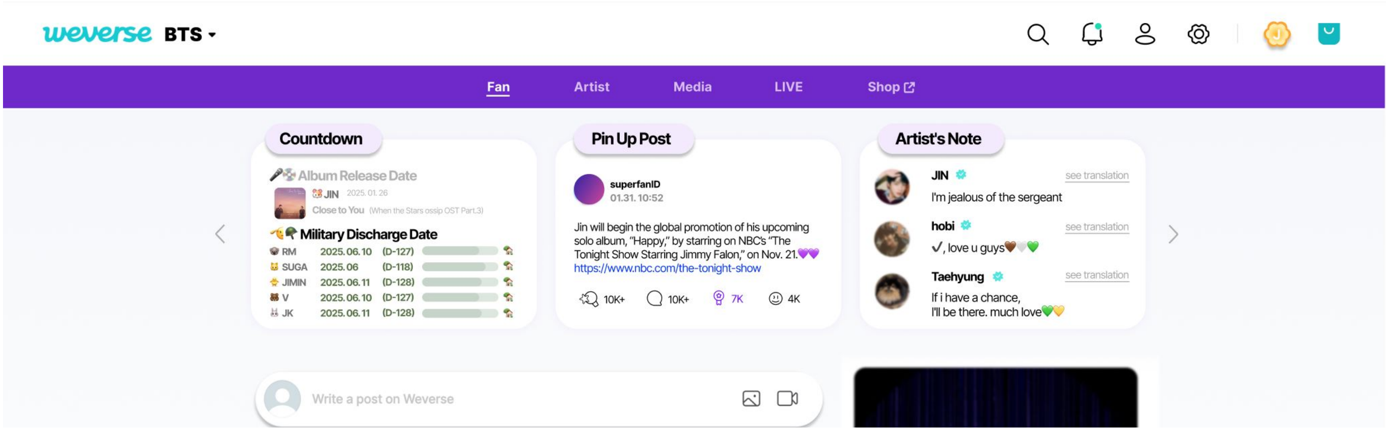


Figure 6. Design Implication for Post-heavy Users: Community Pin-up

cal social connectors who sustain community vitality [61,64]. Extant literature emphasizes that these social interactions serve as the community's "social glue," crucially supporting sustained engagement and collective identity [60,63].

Although comment-heavy users' Authentic scores were similar to those of general users, their consistently higher Tone scores and greater reliance on emojis underline their emotionally expressive communication style. This approach fosters a welcoming and supportive atmosphere, enhancing group solidarity and overall community cohesion [47,61]. Furthermore, frequent emoji usage serves as a critical emotional cue, reinforcing perceived authenticity and emotional connectivity among users—an essential factor in maintaining long-term community engagement and satisfaction [63]. Previous studies have noted that emotional expressiveness, including emoji use and affective language, is strategically employed within fandom communities to deepen social bonds and maintain vibrant participatory cultures [61,64].

Collectively, these insights position comment-heavy users not merely as frequent participants, but as pivotal agents facilitating emotional support, social connectivity, and collective identity formation. Their nuanced communicative strategies and rich emotional engagement significantly contribute to the vitality and sustainability of digital fandom communities, underscoring the importance of their role beyond mere quantitative measures of participation.

6.4. Design Implications

Our findings indicate clear distinctions between post-heavy and comment-heavy users, suggesting that a single, uniform strategy for enhancing digital fandom experiences may not suffice. Instead, distinct, more tailored approaches are recommended. Below, we present potential design implications for improving online fan communities, derived from a one-hour design workshop held on December 19th, 2024, with three service designers, each bringing an average of 7.3 years of service design strategy experience.

6.4.1. For Post-heavy Users: Pinups to Foster a Sense of Community

Post-heavy users often submit near-duplicate posts to maintain visibility in fast-moving feeds. Their posts follow a highly structured, analytic style oriented toward the group, resembling formal statements rather than personal messages. Notably, minimal use of personal pronouns in their posts underscores the communitarian aspect of this user group. Based on the post-heavy user characteristics observed in our study, we propose a design intervention that ensures the visibility of community-driven posts rather than indiscriminately boosting all content. Figure 6 illustrates a design approach aimed at giving more prominence to posts that contribute to shared objectives in fandom communities.

In many digital fandom communities and platforms, a stream-based interface with constant content updates can cause posts to accumulate rapidly, making it difficult for authors to keep their contributions visible. As a result, while a high volume of content is generated, the platform itself may not effectively promote discussion or interaction among fans—something users often expect in these community spaces.

To address this, platforms could be reimagined to better accommodate direct interactions among users and the visibility needs of post-heavy users, thereby moving beyond a predominantly one-way parasocial interaction model. In line with the concepts of "one-and-a-half-way" [68] or even two-way communication [69], a reconfigured layout could cultivate richer community-building. Although two-way engagement is technically possible in many systems, current designs often lean toward a one-way communication format, which constrains deeper interactions.

Redesigning interfaces to support more effective user interactions—including features that highlight (or "pin") fan-made posts—would help foster a sense of community that is pivotal for a thriving fandom. A foundational definition of community [70] outlines four elements: *membership*, *influence*, *integration and fulfillment of needs*, and *shared emotional connection*. Extending this to virtual environments, the concept of *Sense of Virtual Community* (SoVC) [71] includes *recognition of members* and *personal connections*. Here, we emphasize *recognition of members* as particularly valuable for sustaining long-term engagement among post-heavy users. Although *influence* is also relevant, it typically develops over time; likewise, *shared emotional connection* and *ursonal connections* may be more significant for commenting behaviors. Meanwhile, *membership* and *integration and fulfillment of needs* are more collective in nature, lying beyond our focus on individual user actions in digital fandom.

By actively recognizing fan-generated contributions, platforms can motivate users to participate more so their own posts also gain visibility, thereby strengthening SoVC. This approach also bolsters *cohesion*: once community members see high-quality, pinned posts from others, they are more inclined to create related content themselves. Cohesion refers to a sense that the group has well-established norms for activities like chatting and posting, with many users adhering to them [2]. Because cohesion is tied to stronger SoVC and is essential for ongoing participation, these design recommendations differ from typical approaches by applying proven practices from successful fan engagement. Allowing users to share "official," community-wide information on mutual interests—visible to all—can help unite members around shared goals while addressing the visibility concerns of post-heavy users.

The current Weverse platform has a Report and Block user feature that allows users to report spam-like posts or posts that have a negative impact. This can lead to increased surveillance of each other in a community with a large number of users. Community platforms like Reddit utilize trust and reputation systems to leverage the collaborative nature of the community to achieve a natural effect [72]. In this way, the quality of the pin-ups can be maintained based on the fact that the design application reflects user recommendations for duplicate posts, utilizes the surveillance of a large number of users, and allows anyone to report.

6.4.2. For Comment-heavy Users: Contextual Translation and Reaction Matrix

Comment-heavy users tend to be highly open and casual in expressing emotions, frequently introducing custom idioms to strengthen their bonds with fellow fans. A notable example is *Borahae*, which replaces the Korean phrase for "I love you" (*Saranghae*) with the word "purple" (*Bora*, meaning purple in Korean), referring to the signature color of the BTS fandom. Fans also use purple heart emojis (💜) as a symbolic expression of

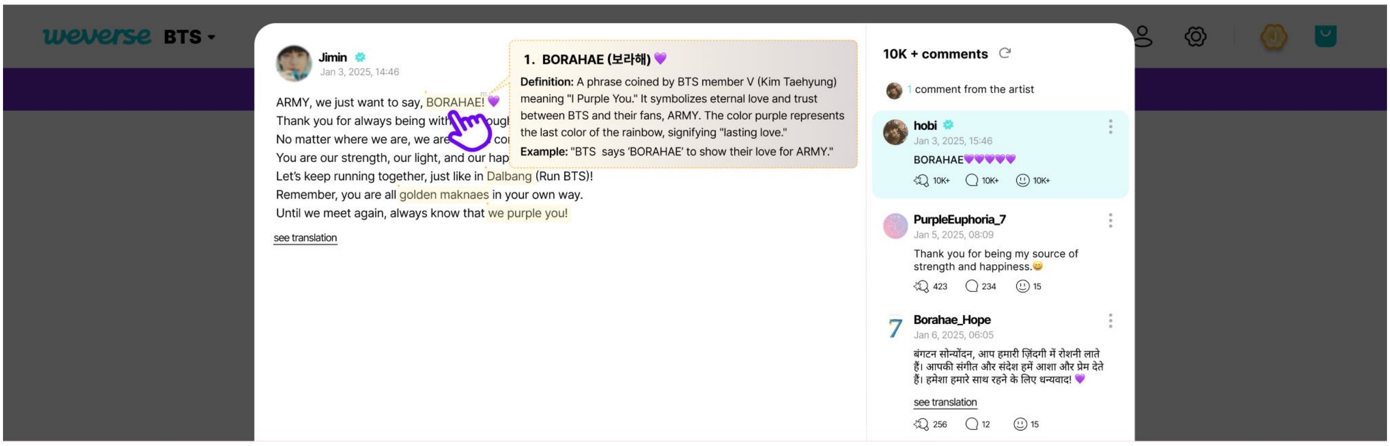


Figure 7. Design Implication for Comment-heavy Users 1: Contextual Translation

affection for BTS (see Figure 5). This phenomenon is not restricted to K-pop: Beyoncé's fans, for instance, often use honey or bee emojis (e.g., 🐝), reflecting her nickname "Queen B," and Ariana Grande's fans have adopted the term "Yuh," derived from her distinctive pronunciation. Emojis, colors, slang, and chants collectively foster a sense of belonging and help shape fandom culture. Prior research on computer-mediated communication suggests that emoticons and emojis can enrich text-based interactions by compensating for the lack of nonverbal cues, thus conveying deeper cultural and emotional context [73–75].

Yet, when these expressions become commonplace among artists or core fandom members, newcomers—lacking familiarity with these references—may struggle to follow and engage with the content. This can diminish their overall experience and satisfaction with fandom services. Moreover, heavy users are typically intensive in both frequency and volume of activity [21,55], and in our study, a small percentage of comment-heavy users produced 80% of the total comments. Consequently, the community atmosphere can become dominated by heavy users' fandom-specific slang and shared memes, making it difficult for newer fans to participate meaningfully.

Further analysis was done to identify and discuss group differences in emoji to deepen our understanding. Among the 100,400 comments posted by 2,270 general users, 48.3% included the purple heart emoji, a widely recognized symbol of fandom culture. In contrast, 32 post-heavy users contributed 7,551 comments, with a significantly higher purple heart usage rate of 71.9%. Similarly, 71 comment-heavy users were responsible for 96,814 comments, with 45.4% containing the purple heart. Although the comment-heavy group consists of only 71 individuals, they contributed approximately 21.5% of the total 200,765 comments. This indicates that a relatively small number of highly active users have a disproportionate influence on the overall comment landscape. These findings highlight the influential role of super fans in shaping the tone and symbolic language of online fandom spaces. However, because such symbolic expressions—like the purple heart emoji—are often developed and reinforced by a small but vocal segment of users, newcomers, who make up a larger proportion of the user base, may struggle to intuitively understand these culturally embedded meanings. This can create subtle communication barriers and hinder inclusive participation within fandom communities. To address this issue, platforms could implement contextual translation features (see Figure 7) that explain key fandom idioms or slang when users hover over or select certain words. In doing so, they can lower the barrier to entry for newer fans and encourage richer feedback, thereby boosting interaction across the fan community.

Comment-heavy users also exhibit higher emotional expressiveness compared to post-heavy users, but many social media–based fan platforms still provide only limited ways

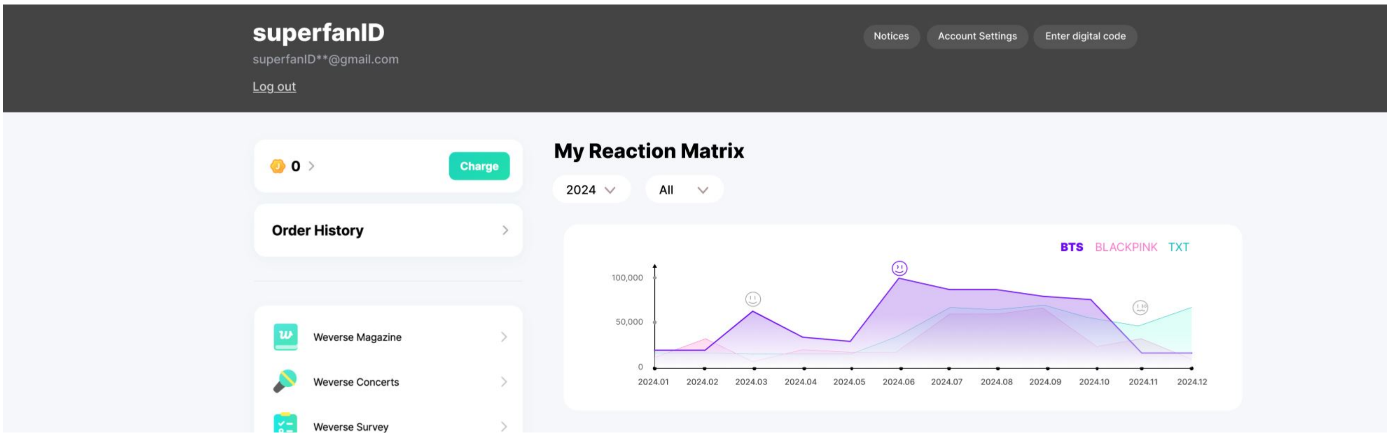


Figure 8. Design Implication for Comment-heavy Users 2: Reaction Matrix

to convey emotion, such as “Like” buttons or text-based replies. Facebook introduced a broader set of reaction buttons (Love, Haha, Wow, Sad, Angry) in 2016, recognizing that a simple “Like” may not always suit posts dealing with serious topics, and that writing a text response may feel burdensome for some users. In many digital fandom contexts, large volumes of posts and comments cluster around specific events, generating intense emotional reactions. Yet these data often go untracked, missing an opportunity for both users and platforms to reflect on[76], and potentially enhance, the fandom experience.

Figure 8 illustrates a “Reaction Matrix” feature, allowing users active in multiple fandoms to observe how their emotional responses shift over time or around certain events. For example, if a user with the ID "superfanID" accessed My Page on a fandom platform, they might see various emotional reactions they have engaged with within different fandoms. As shown in Figure 8, the user is a fan of BTS, Blackpink, and TXT, and has accumulated a year’s worth of emotional reactions related to the certain fandom. This design implication concept captures the dynamics of their emotional expressions over time, for example, showing more joy during periods when a new album was released and increased sadness during times of scandals or controversies. Although this example focuses on fandom love, the same design principle could be adapted for broader emotional insights, reflecting personal informatic approaches that encourage self-awareness and continuous engagement. By helping fans visualize and reflect on their emotional journey, such interfaces have the potential to strengthen community ties while enhancing individual well-being through personal reflection [77]. At a macro level, this design implication could also serve as a new form of audience feedback [78] These impactful engagement features hold significant implications for demand forecasting in the entertainment industry, as they highlight how individuals can subtly shape the content they share.

7. Limitation and Future Work

Although this study provides valuable insights into fan interactions on Weverse, particularly in relation to BTS, and has generalizability across the size and diversity of fandoms, several limitations must be acknowledged. First, our analysis focuses on a single celebrity with a large, global fandom. The behavioral patterns of post-heavy and comment-heavy groups could differ when examining other celebrities, smaller communities, or entirely different fan cultures. This limitation raises questions about the broader generalizability of our findings, suggesting that future work should explore more diverse populations or fandom contexts such as the characteristics of different groups within fandom, such as the gender or nationality of artists. Nevertheless, our core approach—differentiating user groups by posting and commenting activities and analyzing their linguistic behavior—could be readily applied to other fandoms.

We also defined heavy users based on high-level engagement (i.e., posts and comments) rather than lower-level metrics such as views, likes, or follows [25]. Therefore, our study centers on active, public contributors and may underrepresent "lurkers," who read content but do not post or comment [79]. Although our approach implicitly includes some users who rarely engage (for instance, those with zero comments but at least one post), it excludes individuals who never share any content, and we do not know their exact numbers. These lurkers can still influence the community through lower-level engagement such as views, likes, and follows [80]. Future work could explore whether heavy users defined by these alternative engagement types exhibit distinct interaction patterns compared to the post-heavy and comment-heavy groups identified here.

Additionally, our linguistic analysis primarily focused on English-language content because English constituted the largest portion of posts and comments on Weverse (42.4% of posts and 35.8% of comments). Although this approach yielded sufficient data for robust statistical analyses, it inevitably excludes texts in other languages that may reveal important cultural or linguistic nuances. Furthermore, Weverse is an anonymous global platform that collects no demographic information (e.g., age, gender, or nationality), and even English-language posts may be authored by non-native speakers. This introduces variability in language proficiency and style that was not explicitly addressed in our study. Future research could expand upon our findings by investigating content in additional languages, employing careful translation methods to minimize bias, or incorporating demographic factors where available. Such efforts would likely offer deeper insights into the cultural and linguistic dimensions of user engagement in global fandom communities.

Furthermore, in using a k-means clustering approach, we assumed that heavy users would naturally separate into two main categories: post-heavy and comment-heavy. Although the results supported this assumption, k-means is known to be sensitive to outliers. We observed a small number of users who were highly active in both posts and comments; these individuals were nominally assigned to one of the heavy clusters yet represent a potentially distinct subgroup that warrants further investigation.

Although our proposed design implications offer targeted strategies for improving digital fandom experiences, several possibilities merit further exploration. As fandom engagement becomes increasingly influenced by Industry 4.0 technologies, such as AI-powered content recommendations and automated moderation, it is important to understand how AI algorithmic interventions will affect user behavior in their fandom activities. Future studies should examine how AI interventions, such as AI-based personalization and real-time language translation, impact different engagement behaviors within digital fandoms.

Lastly, while our study focused on analyzing activity patterns in a large-scale digital fandom community, it is important to acknowledge ethical considerations surrounding data use and user classification. Although the data we collected consisted only of publicly visible posts and comments that do not include personally identifying information, the act of clustering users based on their behavior can raise broader concerns if applied to user-facing systems. In our case, the clustering was used solely for analytical purposes and not for real-time user monitoring or interface personalization. The design implications we propose are intended to support different styles of engagement at the community level, without disclosing or operationalizing users' cluster identities. Nevertheless, future research should further explore the ethical implications of algorithmic categorization, especially in contexts where such classifications may become visible to users or used to drive platform decisions.

Ensuring transparency, informed consent, and safeguards against algorithmic bias will be essential for the responsible application of user modeling techniques.

Despite the aforementioned limitations, this study makes an important contribution to the HCI community by highlighting how different forms of engagement (posting versus commenting) can reveal unique user behaviors and linguistic styles in a large-scale fandom setting. By refining this methodology and addressing the outlined limitations, researchers and practitioners can apply our approach to a broader range of platforms, user populations, and content domains—ultimately advancing our understanding of community engagement and user behavior in digital environments.

8. Conclusion

In this study, we addressed three research questions concerning digital fandom engagement: (1) how superfans can be classified into subgroups, (2) how their activities differ across user groups, and (3) how linguistic styles vary among these user groups. Using a k means clustering approach based on users' posting and commenting behaviors, we found that a small subset of post heavy and comment heavy superfans contributes a disproportionate share of overall platform activity. Subsequent linguistic analyses revealed that post heavy users often pursue structured, collective goals, frequently posting calls to action or promotional content, whereas comment heavy users tend toward a more casual, socially oriented style, employing colloquial language and heightened emotional expression. These findings highlight the multifaceted nature of heavy engagement in fandom communities, underscoring distinct motivations and communication patterns among subgroups. Building on these insights, we propose design implications such as increasing the visibility of community driven posts and incorporating features like reaction matrices or contextual translation to address each subgroup's needs more effectively. By recognizing and accommodating these diverse user segments, digital fandom platforms can foster sustained engagement, stronger social cohesion, and a richer sense of community.

Author Contributions: Conceptualization, Y.L. and S.P.; methodology, Y.L. and S.P.; Validation, Y.L.; formal analysis, S.P.; Investigation, Y.L.; visualization, S.P.; supervision, S.P.; All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Sungshin Women's University Research Grant of 2021.

Institutional Review Board Statement: This study was exempt from IRB review (exemption number: SSWUIRB-2024-052) because it involved collecting publicly available posts and comments without gathering personally identifiable information such as usernames or email addresses.

Informed Consent Statement: Not applicable.

Data Availability Statement: Our dataset is available at https://doi.org/10.5281/zenodo.14942441 and includes user information (e.g., nicknames, number of posts, number of comments) and LIWC results (e.g., the LIWC analysis for each text) from the Weverse BTS Feed.

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

References

1. Wang, Y.; Lu, Z. Making Sense of Post-match Fan Behaviors in the Online Football Communities. In Proceedings of the Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2023; CHI '23. https://doi.org/10.1145/3544548.3581310.
2. Kairam, S.R.; Mercado, M.C.; Sumner, S.A. A Social-Ecological Approach to Modeling Sense of Virtual Community (SOVC) in Livestreaming Communities. *Proc. ACM Hum.-Comput. Interact.* **2022**, *6*. https://doi.org/10.1145/3555081.
3. Wagenaar, W.F. Discord as a Fandom Platform: Locating a New Playground. *Transformative Works and Cultures* **2024**, *42*. https://doi.org/https://doi.org/10.3983/twc.2024.2473.

4. Macnaughtan, K.; and, V.T. “We will not be lectured”: understanding political fandom on TikTok. *Communication Research and Practice* **2024**, *10*, 148–164, [https://doi.org/10.1080/22041451.2024.2355818]. https://doi.org/10.1080/22041451.2024.2355818.
5. Fiesler, C.; Dym, B. Moving Across Lands: Online Platform Migration in Fandom Communities. *Proc. ACM Hum.-Comput. Interact.* **2020**, *4*. https://doi.org/10.1145/3392847.
6. Kusuma, A.; Purbantina, A.P.; Nahdiyah, V.; Khasanah, U.U. A Virtual Ethnography Study: Fandom and Social Impact in Digital Era. 2020.
7. Ringland, K.E.; Bhattacharya, A.; Weatherwax, K.; Eagle, T.; Wolf, C.T. ARMY’s Magic Shop: Understanding the Collaborative Construction of Playful Places in Online Communities. In Proceedings of the Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2022; CHI ’22. https://doi.org/10.1145/3491102.3517442.
8. Park, S.Y.; Santero, N.K.; Kaneshiro, B.; Lee, J.H. Armed in ARMY: A Case Study of How BTS Fans Successfully Collaborated to #MatchAMillion for Black Lives Matter. In Proceedings of the Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2021; CHI ’21. https://doi.org/10.1145/3411764.3445353.
9. Kang, J.; Lee, M.; Park, E.; Ko, M.; Lee, M.; Han, J. Alliance for My Idol: Analyzing the K-pop Fandom Collaboration Network. In Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2019; CHI EA ’19, p. 1–6. https://doi.org/10.1145/3290607.3312906.
10. Maltby, J.; Day, L.; McCutcheon, L.E.; Houran, J.; Ashe, D. Extreme celebrity worship, fantasy proneness and dissociation: Developing the measurement and understanding of celebrity worship within a clinical personality context. *Personality and Individual Differences* **2006**, *40*, 273–283. https://doi.org/https://doi.org/10.1016/j.paid.2005.07.004.
11. Thorne, S. An exploratory investigation of the theorized levels of consumer fanaticism. *Qualitative Market Research: An International Journal* **2011**, *14*, 160–173.
12. Edlom, J.; Karlsson, J. Keep the Fire Burning: Exploring the Hierarchies of Music Fandom and the Motivations of Superfans. *Media and Communication* **2021**.
13. Fiske, J. *The Adoring Audience, The Cultural Economy of Fandom*, - ed.; Routledge: London, 1992.
14. Duffett, M. *Understanding Fandom: An Introduction to the Study of Media Fan Culture*, - ed.; Bloomsbury Academic: New York, 2013.
15. Abercrombie, N.; Longhurst, B. *Audiences: A sociological theory of performance and imagination*, - ed.; SAGE Publications: California, 1998.
16. Sandvoss, C. *Fans: The Mirror of Consumption*, - ed.; Polity: Cambridge, 2005.
17. Tapp, A.; Clowes, J. From "carefree casuals" to "professional wanderers": Segmentation possibilities for football supporters. *European Journal of Marketing* **2002**, *36*, 1248–1269.
18. Hills, P.; Argyle, M. Positive moods derived from leisure and their relationship to happiness and personality. *Personality and Individual Differences* **1998**, *25*, 523–535. https://doi.org/https://doi.org/10.1016/S0191-8869(98)00082-8.
19. Lee, S.; Lee, J. “Ju. T’aime” My Idol, My Streamer: A Case Study on Fandom Experience as Audiences and Creators of VTuber Concert. *IEEE Access* **2023**, *11*, 31125–31142. https://doi.org/10.1109/ACCESS.2023.3252563.
20. Wang, Y.; Lu, Z. Making Sense of Post-match Fan Behaviors in the Online Football Communities. In Proceedings of the Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2023; CHI ’23. https://doi.org/10.1145/3544548.3581310.
21. Park, S.; Ackerman, M.S.; Lee, U. Localness of location-based knowledge sharing: a study of Naver KiN “Here”. *ACM Transactions on the Web (TWEB)* **2018**, *12*, 1–33.
22. Daphna Shwartz-Asher, Soon Ae Chun, N.R.A. Knowledge behavior model of e-government social media users. *Transforming Government: People, Process and Policy* **2017**, *11*.
23. Huang, Y.T.; Su, S.F. Motives for Instagram Use and Topics of Interest among Young Adults. *Future Internet* **2018**, *10*. https://doi.org/10.3390/fi10080077.
24. Chin, H.; Zhunis, A.; Cha, M. Behaviors and Perceptions of Human-Chatbot Interactions Based on Top Active Users of a Commercial Social Chatbot. *Proc. ACM Hum.-Comput. Interact.* **2024**, *8*. https://doi.org/10.1145/3687022.
25. Aldous, K.K.; An, J.; Jansen, B.J. View, Like, Comment, Post: Analyzing User Engagement by Topic at 4 Levels across 5 Social Media Platforms for 53 News Organizations. In Proceedings of the Proceedings of the International AAAI Conference on Web and Social Media, 2019, Vol. 13, pp. 47–57. https://doi.org/https://doi.org/10.1609/icwsm.v13i01.3208.
26. Kim, M.; Ko, H.; Choi, J.; Lee, J. FlumeRide: Interactive Space Where Artists and Fans Meet-and-Greet Using Video Calls. *IEEE Access* **2023**, *11*, 31594–31606. https://doi.org/10.1109/ACCESS.2023.3249460.
27. Guazzini, A.; Guidi, E.; Cecchini, C.; Milani, M.; Vilone, D.; Meringolo, P. Self-presentation and emotional contagion on Facebook: new experimental measures of profiles’ emotional coherence, 2016, [arXiv:cs.SI/1607.07243].
28. Al-Haddad, S.; Al Khasawneh, M.; Tahboub, M.; Al Ouri, S.T.; Al Arabi, D.A.; Abu Sumaqa, L.; Rihani, A.A. The Effect of Intrinsic and Extrinsic Motivations on Social Media Engagement and Customer Likelihood to Share Content on Facebook. *Int. J. Cyber Behav. Psychol. Learn.* **2022**, *12*, 1–24. https://doi.org/10.4018/IJCBPL.298685.

29. Wang, Y.C.; Dwivedi-Yu, J.; Kraut, R.E.; Halevy, A. Using Comments for Predicting the Affective Response to Social Media Posts . In Proceedings of the 2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII), Los Alamitos, CA, USA, 2023; pp. 1–8. https://doi.org/10.1109/ACII59096.2023.10388133.
30. Wang, X.; Zheng, X.; Zhang, X.; Zeng, K.; Wang, F.Y. Analysis of Cyber Interactive Behaviors Using Artificial Community and Computational Experiments. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **2017**, *47*, 995–1006. https://doi.org/10.1109/TSMC.2016.2615130.
31. Lee, S.; Lee, J. "Ju. T'aime" My Idol, My Streamer: A Case Study on Fandom Experience as Audiences and Creators of VTuber Concert. *IEEE Access* **2023**, *11*, 31125–31142. https://doi.org/10.1109/ACCESS.2023.3252563.
32. Yum, K.; Kim, J. The Influence of Perceived Value, Customer Satisfaction, and Trust on Loyalty in Entertainment Platforms. *Applied Sciences* **2024**, *14*. https://doi.org/10.3390/app14135763.
33. Kim, M.S.; Wang, S.; Kim, S. Effects of Online Fan Community Interactions on Well-Being and Sense of Virtual Community. *Behavioral Sciences* **2023**, *13*. https://doi.org/10.3390/bs13110897.
34. Choi, H.j. Do K-Pop Consumers' Fandom Activities Affect Their Happiness, Listening Intention, and Loyalty? *Behavioral Sciences* **2024**, *14*. https://doi.org/10.3390/bs14121136.
35. Chin, H.; Zhunis, A.; Cha, M. Behaviors and Perceptions of Human-Chatbot Interactions Based on Top Active Users of a Commercial Social Chatbot. *Proc. ACM Hum.-Comput. Interact.* **2024**, *8*. https://doi.org/10.1145/3687022.
36. Panciera, K.; Halfaker, A.; Terveen, L. Wikipedians are born, not made: a study of power editors on Wikipedia. In Proceedings of the Proceedings of the 2009 ACM International Conference on Supporting Group Work, New York, NY, USA, 2009; GROUP '09, p. 51–60. https://doi.org/10.1145/1531674.1531682.
37. Quattrone, G.; Capra, L.; De Meo, P. There's No Such Thing as the Perfect Map: Quantifying Bias in Spatial Crowd-sourcing Datasets. In Proceedings of the Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, New York, NY, USA, 2015; CSCW '15, p. 1021–1032. https://doi.org/10.1145/2675133.2675235.
38. Tausczik, Y.R.; Pennebaker, J.W. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* **2010**, *29*, 24–54. https://doi.org/10.1177/0261927X09351676.
39. Boyd, R.L.; Ashokkumar, A.; Seraj, S.; Pennebaker, J.W. The development and psychometric properties of LIWC-22. Austin, TX: University of Texas at Austin, 2022.
40. Pennebaker, J.W.; Mehl, M.R.; Niederhoffer, K.G. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology* **2003**, *54*, 547–577.
41. Saha, K.; Sharma, A. Causal Factors of Effective Psychosocial Outcomes in Online Mental Health Communities. *Proceedings of the International AAAI Conference on Web and Social Media* **2020**, *14*, 590–601. https://doi.org/10.1609/icwsm.v14i1.7326.
42. Jiang, S.; Wilson, C. Linguistic Signals under Misinformation and Fact-Checking: Evidence from User Comments on Social Media. *Proc. ACM Hum.-Comput. Interact.* **2018**, *2*. https://doi.org/10.1145/3274351.
43. Kline, R.B. *Becoming a behavioral science researcher: A guide to producing research that matters*; Guilford Press, 2008.
44. Pennebaker, J.W.; Chung, C.K.; Frazee, J.; Lavergne, G.M.; Beaver, D.I. When Small Words Foretell Academic Success: The Case of College Admissions Essays. *PLOS ONE* **2014**, *9*, 1–10. https://doi.org/10.1371/journal.pone.0115844.
45. Boyd, R.L.; Pennebaker, J.W. Did Shakespeare Write Double Falsehood? Identifying Individuals by Creating Psychological Signatures With Text Analysis. *Psychological Science* **2015**, *26*, 570–582. https://doi.org/10.1177/0956797614566658.
46. Jordan, K.N.; Sterling, J.; Pennebaker, J.W.; Boyd, R.L. Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proceedings of the National Academy of Sciences* **2019**, *116*, 3476–3481. https://doi.org/10.1073/pnas.1811987116.
47. Fox, A.K.; Stafford, M.B.R. Olympians on Twitter: A Linguistic Perspective of the Role of Authenticity, Clout, and Expertise in Social Media Advertising. *Journal of Current Issues & Research in Advertising* **2020**, *42*, 294 – 309.
48. Michelle Drouin, Ryan L. Boyd, J.T.H.; James, A. Linguistic analysis of chat transcripts from child predator undercover sex stings. *The Journal of Forensic Psychiatry & Psychology* **2017**, *28*, 437–457. https://doi.org/10.1080/14789949.2017.1291707.
49. Kacewicz, E.; Pennebaker, J.W.; Davis, M.; Jeon, M.; Graesser, A.C. Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology* **2014**, *33*, 125–143. https://doi.org/10.1177/0261927X13502654.
50. Newman, M.L.; Pennebaker, J.W.; Berry, D.S.; Richards, J.M. Lying Words: Predicting Deception from Linguistic Styles. *Personality and Social Psychology Bulletin* **2003**, *29*, 665–675. https://doi.org/10.1177/0146167203029005010.
51. Markowitz, D.M.; Kouchaki, M.; Gino, F.; Hancock, J.T.; Boyd, R.L. Authentic First Impressions Relate to Interpersonal, Social, and Entrepreneurial Success. *Social Psychological and Personality Science* **2023**, *14*, 107–116. https://doi.org/10.1177/19485506221086138.
52. Monzani, D.; Vergani, L.; Pizzoli, S.F.M.; Marton, G.; Pravettoni, G. Emotional Tone, Analytical Thinking, and Somatosensory Processes of a Sample of Italian Tweets During the First Phases of the COVID-19 Pandemic: Observational Study. *J Med Internet Res* **2021**, *23*, e29820. https://doi.org/10.2196/29820.

53. Cohn, M.A.; Mehl, M.R.; Pennebaker, J.W. Linguistic Markers of Psychological Change Surrounding September 11, 2001. *Psychological Science* **2004**, *15*, 687–693. https://doi.org/10.1111/j.0956-7976.2004.00741.x.
54. Suh, B.; Convertino, G.; Chi, E.H.; Pirolli, P. The singularity is not near: slowing growth of Wikipedia. In Proceedings of the Proceedings of the 5th International Symposium on Wikis and Open Collaboration, New York, NY, USA, 2009; WikiSym '09. https://doi.org/10.1145/1641309.1641322.
55. Wang, G.; Gill, K.; Mohanlal, M.; Zheng, H.; Zhao, B.Y. Wisdom in the social crowd: an analysis of quora. In Proceedings of the Proceedings of the 22nd International Conference on World Wide Web, New York, NY, USA, 2013; WWW '13, p. 1341–1352. https://doi.org/10.1145/2488388.2488506.
56. Ferrara, E. Twitter spam and false accounts prevalence, detection, and characterization: A survey. *First Monday* **2022**, *27*. https://doi.org/10.5210/fm.v27i12.12872.
57. Qayyum, H.; Zi Hao Zhao, B.; Wood, I.; Ikram, M.; Kourtellis, N.; Ali Kaafar, M. A longitudinal study of the top 1% toxic Twitter profiles. In Proceedings of the Proceedings of the 15th ACM Web Science Conference 2023, New York, NY, USA, 2023; WebSci '23, p. 292–303. https://doi.org/10.1145/3578503.3583619.
58. Sultana, B.; Hridoy, M.N.H.; Islam, M.M.; Amin, R.; Rahman, F. Restriction Algorithm for Duplicate Posts on Online Social Networks: Facebook. *GUB Journal of Science and Engineering* **2020**, *6*, 33–38. https://doi.org/10.3329/gubjse.v6i1.52048.
59. Chu, Z.; Gianvecchio, S.; Wang, H.; Jajodia, S. Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg? *IEEE Transactions on Dependable and Secure Computing* **2012**, *9*, 811–824. https://doi.org/10.1109/TDSC.2012.75.
60. Lee, S.; Tak, J.; Kwak, E.; Lim, T. Fandom, Social Media, and Identity Work: The Emergence of Virtual Community through the Pronoun "We". *Psychology of Popular Media* **2020**, *9*, 436–446. Publisher Copyright: © 2019 American Psychological Association, https://doi.org/10.1037/ppm0000259.
61. Phan, T.T.; Michoud, C.; Volpato, L.; Carral, M.D.R.; Gatica-Perez, D. Health Talk: Understanding Practices of Popular Professional YouTubers. In Proceedings of the Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia, New York, NY, USA, 2022; MUM '22, p. 218–230. https://doi.org/10.1145/3568444.3568453.
62. Wang, Y.; Lv, Q. Jump off the Bandwagon? Characterizing Bandwagon Fans' Future Loyalty in Online NBA Fan Communities . In Proceedings of the 2023 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Los Alamitos, CA, USA, 12 2023; pp. 953–960. https://doi.org/10.1109/ISPA-BDCloud-SocialCom-SustainCom59178.2023.00158.
63. Yan, Q.; Yang, F. From parasocial to parakin: Co-creating idols on social media. *New Media & Society* **2021**, *23*, 2593–2615. https://doi.org/10.1177/1461444820933313.
64. Achilles, L.; Mandl, T.; Womser-Hacker, C. Body like an idol: K-pop fitspiration on Tumblr – an analysis of texts and images. *Finnish Journal of eHealth and eWelfare* **2023**, *15*, 380–394. https://doi.org/10.23996/fjhw.126897.
65. Marco, N.D.; Loru, E.; Bonetti, A.; Serra, A.O.G.; Cinelli, M.; Quattrociocchi, W. Patterns of linguistic simplification on social media platforms over time. *Proceedings of the National Academy of Sciences* **2024**, *121*, e2412105121. https://doi.org/10.1073/pnas.2412105121.
66. Zappavigna, M.; Logi, L. How emoji make meaning and enact ambient affiliation: a social semiotic account of emoji-text relations in TikTok comments. *Social Semiotics* **2024**, *0*, 1–21. https://doi.org/10.1080/10350330.2024.2389520.
67. Chmiel, A.; Sienkiewicz, J.; Thelwall, M.; Paltoglou, G.; Buckley, K.; Kappas, A.; Hołyst, J.A. Collective Emotions Online and Their Influence on Community Life. *PLOS ONE* **2011**, *6*, 1–8. https://doi.org/10.1371/journal.pone.0022207.
68. Kowert, R.; Daniel, E. The one-and-a-half sided parasocial relationship: The curious case of live streaming. *Computers in Human Behavior Reports* **2021**, *4*, 100150. https://doi.org/https://doi.org/10.1016/j.chbr.2021.100150.
69. Marwick, A.; danah boyd. To See and Be Seen: Celebrity Practice on Twitter. *Convergence* **2011**, *17*, 139–158. https://doi.org/10.1177/1354856510394539.
70. McMillan, D.; Chavis, D. Sense of community: A definition and theory. *Journal of Community psychology* **1986**, *14*, 6–23. https://doi.org/10.1002/1520-6629(198601)14:1<6::AID-JCOP2290140103>3.0.CO;2-I.
71. Blanchard, A.L.; Markus, M.L. The experienced "sense" of a virtual community: characteristics and processes. *SIGMIS Database* **2004**, *35*, 64–79. https://doi.org/10.1145/968464.968470.
72. Jøsang, A.; Ismail, R.; Boyd, C. A survey of trust and reputation systems for online service provision. *Decision Support Systems* **2007**, *43*, 618–644. Emerging Issues in Collaborative Commerce, https://doi.org/https://doi.org/10.1016/j.dss.2005.05.019.
73. Wang, Y.; Li, Y.; Gui, X.; Kou, Y.; Liu, F. Culturally-Embedded Visual Literacy: A Study of Impression Management via Emoticon, Emoji, Sticker, and Meme on Social Media in China. *Proc. ACM Hum.-Comput. Interact.* **2019**, *3*. https://doi.org/10.1145/3359170.
74. Bae, G.; Kwak, D.; Lim, Y.k. Why I Choose This Sticker When Chatting with You: Exploring Design Considerations for Sticker Recommendation Services in Mobile Instant Messengers. *Proc. ACM Hum.-Comput. Interact.* **2024**, *8*. https://doi.org/10.1145/3687063.
75. Cha, Y.; Kim, J.; Park, S.; Yi, M.Y.; Lee, U. Complex and Ambiguous: Understanding Sticker Misinterpretations in Instant Messaging. *Proc. ACM Hum.-Comput. Interact.* **2018**, *2*. https://doi.org/10.1145/3274299.

76. Bae, J.e.; Lim, Y.k.; Bang, J.b.; Kim, M.s. Ripening room: designing social media for self-reflection in self-expression. In Proceedings of the Proceedings of the 2014 Conference on Designing Interactive Systems, New York, NY, USA, 2014; DIS '14, p. 103–112. https://doi.org/10.1145/2598510.2598567.
77. Cantrell, S.J.; Winters, R.M.; Kaini, P.; Walker, B.N. Sonification of Emotion in Social Media: Affect and Accessibility in Facebook Reactions. *Proc. ACM Hum.-Comput. Interact.* **2022**, *6*. https://doi.org/10.1145/3512966.
78. Erickson, J.; Yan, B. Affective Design: The Influence of Facebook Reactions on the Emotional Expression of the 114th US Congress. In Proceedings of the Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2024; CHI '24. https://doi.org/10.1145/3613904.3641935.
79. Edelmann, N. Reviewing the Definitions of "Lurkers" and Some Implications for Online Research. *Cyberpsychology, Behavior, and Social Networking* **2013**, *16*, 645–649. https://doi.org/10.1089/cyber.2012.0362.
80. Takahashi, M.; Fujimoto, M.; Yamasaki, N. The active lurker: influence of an in-house online community on its outside environment. In Proceedings of the Proceedings of the 2003 ACM International Conference on Supporting Group Work, New York, NY, USA, 2003; GROUP '03, p. 1–10. https://doi.org/10.1145/958160.958162.