first monday
PEER-REVIEWED JOURNAL ON THE INTERNET

Pundits, presenters, and promoters: Investigating gaps in digital production among social media users using self-reported and behavioral measures

by Jiang Ke, Lance Porter, Rui Wang, Seon-Woo Kim, and Martin Johnson

Abstract

Through the lens of Bourdieu's field theory, we investigate the relationship between the social characteristics of social media users and their differentiating practices in producing digital content. Matching survey data with self-reported user profiles and one year of actual posts on Twitter, we found four online fields of lifecasting, politics, promotion, and entertainment. Users tweeting positively about entertainment held higher levels of social capital. From 2011 and 2017, we found a reduction in lifecasting was accompanied by the rise of promotion.

Contents

Introduction

For the better part of the last three decades, as we have steadily moved to a mostly digital distribution model for media content, scholars have expressed concern about unequal access to economic, political, social, and cultural content distributed via the internet. Accordingly, scholars have studied the "digital divide" or "the gap between those who do and those who do not have access to new forms of information technology" [1]. Digital inequalities tend to mirror the off-line areas of inequality such as race and ethnicity (Mesch and Talmud, 2011), gender (Ono and Zavodny, 2008), and socioeconomic status (Witte and Mannon, 2010).

Scholars initially were concerned about the effects of differences in access to new information and communication technologies among different populations. Bill Clinton and Al Gore first spoke of the digital divide in 1996 as an access issue between those of higher means and those of lower socioeconomic status (Schradie, 2011). However, scholars have pointed out that simple access does not equal use. Over time, researchers expanded the concept of the digital divide to a second level (Riggins and Dewan, 2005). The first level refers to the aforementioned inequality of access to information and communication technologies (ICTs). The second level refers to use or the digital capability divide (Riggins and Dewan, 2005; Robinson, *et al.*, 2015), which refers to the ability to use the Internet (Harris, *et al.*, 2017). The digital capability divides result from difficulties in access and other factors such as skills, participation, and education. The second level-divide likely results from structural considerations as well, such as the characteristics of social environments expressed on social media and in other online communication networks. Longitudinal studies have shown over time that as the first-level digital divide has decreased in some Western countries, the second-level digital divide has become more entrenched (Helsper and Reisdorf, 2017). Additionally, other motivational reasons for non-use have arisen among this small but pervasive "digital underclass" such as a lack of interest in ICTs [2]. In fact, attitudes toward ICTs have been found to play as much of a role in ICT uses as socioeconomic factors (Reisdorf and Groselj, 2017).

Digital scholars have found Pierre Bourdieu's interrelated concepts of field, capital, and habitus useful in explaining digital inequality (Gilbert, 2010; Halford and Savage, 2010; Ignatow and Robinson, 2017; Levina and Arriaga, 2014; Lutz, 2016; Reiss and Tsvetkova, 2020; Sims, 2014; Yates and Lockley, 2018). According to Bourdieu (2004, 1990, 1984), a field can be defined as a network of interacting social agents, the habitus of whom represents the transportation of objective structures of the field into subjective structures of action and thoughts. The interrelations and power within the field are determined by the capital that is the accumulated labor enabling the social agents in the field to appropriate social energy in the form of reified or living labor (Bourdieu, 1986).

Lindell (2017) applied field theory to the field of culture in Sweden by examining Facebook likes between cultural institutions. Similarly, we use those concepts here to investigate one dimension of this digital capability divide — differences in digital content production — using a novel research design that allows us to explore both self-reported patterns of social media expression and online behavior expressed on Twitter. According to surveys conducted by the Pew Research Center, 22 percent of Americans use Twitter (Pew Research Center, 2021) and 71 percent of those users seek news on Twitter (Odabaş, 2022). Furthermore, Phua, *et al*. (2017) found that Twitter users tended to have the most social capital among users of other social media platforms such as Facebook, Instagram, and Snapchat. Twitter also remains a mostly public source of social expression. These expressions are mostly produced by a small group of active users, with 80 percent of Twitter content produced by 10 percent of users (Wojcik and Hughes, 2019). While Twitter users interact in social networks, these are, in general, not closed communities. Some users choose to protect their Twitter content from public viewing, but most open their tweets to the entire Internet. In turn, this has attracted the attention of researchers hoping to use social media expression as a surrogate for public opinion (O'Connor, *et al*., 2010).

We collected substantive survey responses from a panel of Twitter users, including their Twitter account names. Using a Python script, we archived seven years of public Twitter data from these users using the Twitter application programming interface (API). Using this data, we developed topic models to examine the dominant fields of Twitter, as well as the evolution of those fields. We also matched the survey data completed in 2016 with self-reported user profiles and a year's worth of actual posts on Twitter in 2016 to examine the differences in the online fields, capital, and habitus among different racial and socioeconomic groups.

Background and research questions

The digital divide and online content creation

Certainly, one of the most important second-level disparities is the effect of the digital divide on the production of content. One of the promises of the digital revolution is that the internet disrupts existing media power structures, enabling citizens in online content creation (Benkler and Nissenbaum, 2006; Brake, 2014; Bruns, 2008; Jenkins, 2006). Particularly with the rise of social media and Web 2.0 platforms such as Twitter and Facebook, scholars have conceptualized these production capabilities as prosumption (Ritzer and Jurgenson, 2010), produsage (Bruns, 2008), pro/am creation (Leadbeater, 2008), digital production (Schradie, 2011), peer production (Benkler and Nissenbaum, 2006), practice (Sims, 2014) and digital participation (Hoffmann, *et al*., 2015). However, the digital divide is even wider at the second level in terms of who participates in the creation of content (Robinson, *et al*., 2015). The COVID-19 pandemic seems to be exacerbating the situation as students with fewer digital skills are facing increased difficulty participating in the new digital learning environment (Ceres, 2020). Scholars have studied this particular aspect of the digital divide as the participation divide (Robles Morales, *et al*., 2016), digital production gap (Schradie, 2011), digital engagement gap (Robinson, *et al*., 2015) and participation gap (Jenkins, *et al*., 2009).

Digital inequality researchers have used Bourdieu's field theory to examine the status differences in online usage in

Europe (Lindell, 2017; Lutz, 2016; Mihelj, *et al.*, 2019; Reiss and Tsvetkova, 2020; Zillien and Marr, 2013), U.S. (Hargittai and Hinnant, 2008), and Israel (Arie and Mesch, 2015). According to Bourdieu's (1984) field theory, society is made up of overlapping fields, each having its own rules and logic whereby actors in these fields struggle and compete to accumulate different forms of capital within the fields. More successful actors can then reinvest different types of capital to obtain and maintain success within the fields. Lindell (2017) called for researchers to move beyond self-reported data to examine "objective relations' and discover "crystalized acts of recognition" [3]. Levina and Arriaga (2014) proposed the concept of an online field to examine power relations online in relation to user generated content. Here we use both behavioral and survey data to determine online fields and actors and the demographic differences between them. Therefore, we propose the following research questions:

> *RQ1*: What are the online fields of Twitter?
> *RQ2*: How do these fields evolve over time in terms of contents and Twitter users?

Bourdieu's central concept of the habitus refers to the internalization of a field as exhibited by the relationships between certain logics, schemata, or ideology. People develop habitus by interacting within fields, which produces patterns of behavior that are attempts at gaining capital within those fields. Individuals develop schemata such as likes, accents, postures, expressiveness, and linguistic practices through daily social interactions in and around fields. By examining habitus, digital inequality researchers have shown how the use of information technologies is mediated by the access to economic resources. For example, Robinson (2018, 2009) found that highly connected upper-middle income youths develop a serious play habitus, while disadvantaged youths enact a more task-oriented approach to information technology with frequent interruptions to their connectivity. These differences in use create a gap where upper-middle income youths develop superior media skills, which allows them to benefit from the technology more than disadvantaged youths. To examine digital inequality, Ignatow and Robinson (2017) also measured aspects of habitus by coding interactions between sentiment, emotions, tone, and volume of expression.

Extending the principal forms of capital (*i.e.*, cultural, symbolic, economic, and social capital) from Bourdieu, digital capital refers to "the reach, scale and sophistication of his or her online behavior" [4]. Digital capital has been studied by examining the digital footprints left behind by social media platforms such as Twitter and Facebook (Hofer and Aubert, 2013; Lewis, *et al.*, 2008). Digital capital is also considered secondary to the more primary forms of capital such as economic and cultural (Ignatow and Robinson, 2017). Social media activity more often leads to the social form of digital capital, rather than more skills-based digital capital such as programming knowledge, which can more directly lead to economic capital through labor skills (Ignatow and Robinson, 2017; Levina and Arriaga, 2014). Successful digital producers whose content is shared and favorited the most can then earn digital capital by capturing the attention of consumers on a platform. Our work examines the relationships between online fields, habitus, and the social form of digital capital by tracing actors' tweets and matching their Twitter behavior with the self-reported survey data. Therefore, we propose the following research questions:

> *RQ3*: What is the habitus of actors participating in the different online fields of Twitter?
> *RQ4*: How is social capital accumulated by different types of actors across different fields?

Methodology

Sampling

We built a novel social media panel of 3,811 survey respondents with Twitter accounts, identified by Qualtrics Online Panels. We invited these respondents to participate in an online survey fielded 10 June–28 July 2016, for compensation. Of these, 904 accepted additional compensation to provide us with a verified Twitter account name. We required consenting survey participants to log in to their Twitter account from our survey instrument, which verified their Twitter account. The participants also consented to have their tweets mined and analyzed. Among the 904 survey respondents who provided us with verified Twitter account names, 634 respondents' Twitter accounts were public. We thus analyzed these 634 accounts. The respondents were 67.4 percent female, 68.6 percent

Caucasian, 50.9 percent with a college degree, 19.6 percent Republican, 52.4 percent Democrat, and 25.2 percent Independent, with an average age of 39.6 years (*SD* = 12.0). To protect the participants' identities, this paper mainly presents results from aggregate data, and we also present a few paraphrased quotes (removing ID handles) to reflect topics that emerged (Townsend and Wallace, 2016; Williams, *et al*., 2017). We did not examine the content of direct messages, only the total number of times users sent those messages.

Measures

We asked survey participants a series of demographic, political, social, and media-related questions. The independent variables in the present study included family income, education, age (in years), ethnicity (dummy variables for African American, Latinx and other non-white participants, with white as the excluded category), and gender.

We weighted all our analyses to be consistent with the estimated population profile of U.S. Twitter users as reported by the Pew Research Center. Consequently, we computed sampling weights as a function of gender, age, family income, education, and race/ethnicity. These weights helped correct for biases in our sample of Twitter users. Using the Twython package in Python (McGrath, 2013), we collected 634 public Twitter users' 680,560 English tweets from 2011 to 2017 and matched the survey data with social media data in 2016.

To answer *RQ1* (What are the online fields of Twitter?), we conducted a topic modeling analysis of the content of English-language tweets produced by 634 public participants. Specifically, we used Latent Dirichlet Allocation, LDA (Blei, *et al*., 2003), which is one of the most widely used topic modeling algorithms. Before topic modeling, we preprocessed text by removing stopwords and stemming. Using ConText software (Diesner, *et al*., 2013), we examined a range of topic *k* from 3 to 20, extracting corresponding top 100-word lists for each. Then, four Ph.D. researchers from mass communication and political science reviewed each group of topics and selected the topic number based not only on the topic coherence score that reflected the cohesive meaning within topics but also the mutually exclusive meaning between topics. LDA also models each tweet as a mixture of topics. It examines the per-tweet-per-topic probabilities, called γ, ranging from 0 to 1. The value 0 indicates that no words in the tweet were generated from the topic, and the value 1 means all words in the tweet were corresponding to the topic. Using the ConText software, we calculated the γ of each topic for each tweet posted in 2016, and then aggregated the weight of each topic for each participant using the mean value of γ. We also compared the weight of topics identified in *RQ1* among different demographic groups (*i.e*., gender, race, ethnicity, age, education, and income).

To answer *RQ2* (How do these fields evolve over time in terms of contents and Twitter users?), we calculated the yearly average topic proportions of each field. In addition, we extracted 93 active users from 634 respondents when they had tweeted at least once per year from 2011 to 2017 to figure out how content tweeted by active users was different from those of overall users. We conducted a repeated measures ANOVA with a Greenhouse-Geisser correction to examine the evolution of Twitter conversations.

Regarding *RQ3* (What are the habitus of actors participating in different online fields of Twitter?), we first identified various types of actors participating in the Twitter online fields by conducting a K-mean cluster analysis based on the z-scores of the weights of each topic of each participant with kmeans function from the cluster package in R (Maechler, *et al*., 2021). The distance matrix was euclidean. To decide the number of cluster K, we assigned the K value from 3 to 10 for the K-means cluster analysis and calculated the respective inertias that reflect how well a dataset was clustered by K-means by measuring the sum of squares of all dataset points to their closet centroid. Although the value of inertia decreased as the number of clusters increased, we used the model for K-means with low inertia and a low number of clusters (Amelia, 2018). We eventually assigned the K value as 3 since the change in the value of inertia was not significant when K was greater than 3. We then analyzed the habitus of the three different types of actors by looking into the behavioral patterns demonstrated from the results of the k-mean cluster analysis and compared the demographic differences among different types of actors.

We also followed Ignatow and Robinson's (2017) measurement of habitus in their study of digital inequality to compare the sentiment and emotions between the three types of actors and different demographic groups. Specifically, we used polarity and subjectivity as two indicators of sentiment and subjective emotion expressed respectively on Twitter. The polarity and subjectivity classifiers are the two most common lexicon-based strategies for Twitter sentiment analysis (Kharde and Sonawane, 2016; Yaqub, *et al*., 2018). Subjectivity refers to the classification of sentences as subjective opinions or objective facts by using a dictionary to quantify the opinion words (*e.g*., adjectives, adverbs, group of verbs and nouns) (Kharde and Sonawane, 2016). The value of subjectivity ranged from 0 to 1. A value close to 0 indicates an objective tweet, while a value close to 1 indicates a highly subjective tweet. Polarity refers to whether the expressed opinion in a tweet is positive, negative, or neutral by using a sentiment dictionary to assign sentiment scores to the opinion words identified by an analysis of subjectivity (Kharde and Sonawane, 2016). The polarity scores range from -1 to 1, with -1 being most negative and 1 most positive. A polarity score of 0 indicates a neutral sentiment. We used the TextBlob Sentiment library and Natural Language ToolKit (Shah, 2020) in Python to compute the value of polarity and subjectivity for each tweet.

To answer *RQ4* (How is social capital accumulated by different types of actors across different fields?), we measured social capital by examining user influence on Twitter, which can be determined by many factors including not only their production of content but also user engagement with that content (Jain and Sinha, 2020). In this study, we analyzed the production of content by investigating the number of tweets, retweets, direct messages, photos, and videos posted in 2016. The ratio of the number of retweets to the frequency of posts (RRP) was calculated as an indicator of Twitter users' original content production. A lower value indicated a higher production of original content. We also measured user engagement using the number of followers, the number of times users' tweets were retweeted and favorited, and the number of times users were listed as members of Twitter "lists" by other users. We compared the differences of these variables as indicators of social capital among not only different types of actors but also different demographic groups. Furthermore, using the lavaan package in R (Rosseel, 2012), we tested a path model with maximum likelihood estimation to explore the causal relations between actors' self-reported user profiles, online fields of Twitter, and actors' social capital. Political ideology and news consumption variables from survey responses were also added into the modeling. Social capital in the path model was defined as the log of the sum of the normalized value of both production of content and user engagement.

Results

For *RQ1* (What are the online fields of Twitter?), we selected four topics and named them "lifecasting," "promotion," "politics," and "entertainment." These topics correspond to the online fields of Twitter. "Lifecasting" refers to people tweeting about their personal lives, such as their work, family, children, and friends. A fictitious example based on a real tweet in lifecasting is "Thank you all for your 24th birthday wishes. It's wonderful to be celebrated by my family and friends." In the "promotion" field, people tweeted to share coupons or information related to promotional contests. For example, an imaginary tweet is "Apply for this new card. You can get 60,000 points as a sign-up bonus." Under the "politics" field, people tweeted about issues related to the president and political events, such as the U.S. presidential election. For example, a exemplar user may tweet, "I feel somewhat concerned, but the mail-in ballot works well during this pandemic." Under the "entertainment" field, people tweeted about games, reality shows, movies, sporting events, and music. A paraphrased example entertainment tweet would read, "The first episode of this new drama is extremely boring!" Table 1 lists the five most frequent words used in the four fields from 2011 to 2017. Appendix B compares the production of content around four fields between demographic groups.

	2011	2012	2013	2014	2015	2016	2017
LC	**day**	**day**	**day**	**day**	**day**	**day**	**day**
	880	1332	1489	1852	2235	3314	2701
	think	**think**	**think**	**think**	**think**	**think**	**think**
	525	783	840	905	1394	2546	1514
	work	**work**	**life**	**work**	**work**	**friend**	**life**
	412	637	651	827	1195	1535	1406
	feel	**feel**	**feel**	**feel**	**life**	**life**	**work**
	339	578	617	792	894	1436	1220
	life	**life**	**work**	**life**	**family**	**work**	**kid**
	247	467	592	773	802	1391	1152
PR	**free**	**win**	**win**	**win**	**win**	**win**	**giveaway**
	467	1610	3622	4271	8239	14386	26690
	win	**free**	**giveaway**	**free**	**earn**	**earn**	**win**
	459	727	1743	1901	3140	8229	25648
	giftcard	**reward**	**free**	**giveaway**	**free**	**mplusreward**	**earn**
	192	686	1664	1693	3070	7217	7209
	earn	**viggleget**	**giftcard**	**giftcard**	**sywsweep**	**free**	**mplusreward**
	147	596	631	1445	2439	5388	5751
	swagbuck	**giveaway**	**earn**	**earn**	**giftcard**	**giveaway**	**free**
	86	382	630	1146	1686	3917	5528
PO	**obama**	**obama**	**syria**	**obama**	**trump**	**trump**	**trump**
	29	209	135	27	121	1597	2379
	trump	**syria**	**obama**	**syria**	**obama**	**clinton**	**obama**
	8	153	57	10	119	354	179
	syria	**romney**	**clinton**	**trump**	**clinton**	**obama**	**clinton**
	6	144	6	3	44	337	88
		clinton	**trump**	**romney**	**syria**	**syria**	**syria**
		26	6	2	15	26	27
		trump	**romney**	**clinton**		**romney**	**romney**
		26	1	1		9	3
EN	**watch**	**watch**	**watch**	**video**	**gameinsight**	**Play**	**video**
	440	2106	1608	1513	4601	3457	3905
	video	**blogtalkradio**	**play**	**watch**	**video**	**Video**	**play**
	345	855	616	949	3165	2183	3778
	play	**listen**	**game**	**game**	**play**	**Game**	**watch**
	203	746	610	704	2020	1888	2195
	game	**play**	**video**	**play**	**game**	**Watch**	**game**
	171	436	550	623	1445	1489	1631
	song	**nickmom**	**listen**	**listen**	**watch**	**Music**	**playlist**
	157	429	529	397	912	1380	1289

Lc: Lifecasting; PR: Promotion; PO: Politics; EN: Entertainment

Table 1: Five most frequent words used in the four fields from 2011 to 2017.

Lifecasting

We found race and age had a significant impact on the weight of lifecasting. African Americans tweeted more about their personal lives than Non-African Americans; Whites posted significantly less about their personal lives than non-whites. Age had significant negative correlations (r = -.2, p < .000) with the weight of lifecasting, indicating younger people tweeted more about their daily personal lives. In terms of age, users in their 20s tweeted the most about their personal lives, and users in their 50s tweeted the least on this topic.

Politics

Gender, income, and age are significant factors influencing the weight of the field of politics. Men and users whose incomes range from US$50,000 to US$74,999, tweeted significantly more than other women and those at other income levels on politics. Also, age was positively associated with the weight of politics (r = .1, p = .018), indicating older users tweeted slightly more about politics. In terms of age, users in their 60s tweeted the most about politics, and users in their 30s tweeted the least about politics.

Promotion

Gender, age, and education levels are significant factors predicting the weight of promotion. Women tweeted more about promotions than men. While the weight of promotion correlated positively with age ($r = .14, p = .001$), promotion correlated negatively with education ($r = -.17, p < .000$). This indicates older users and lower-educated people tweeted significantly more about promotion. In terms of age, while users in their 50s tweeted significantly more than other age groups about promotion, users in their 20s tweeted the least in this field. In terms of education levels, while people whose degrees were high school or less tweeted the most about promotion, people with postgraduate degrees (*e.g.*, M.A., M.S., Ph.D., M.D., J.D.) tweeted the least about promotion.

Entertainment

There were no significant differences in tweeting about entertainment between different race, gender, age, education, and income groups.

To answer *RQ2* (How did the online fields of Twitter conversation evolve over time?), Figure 1 demonstrates the evolution of the yearly average weight of the four fields among all 634 public users and 93 active users.

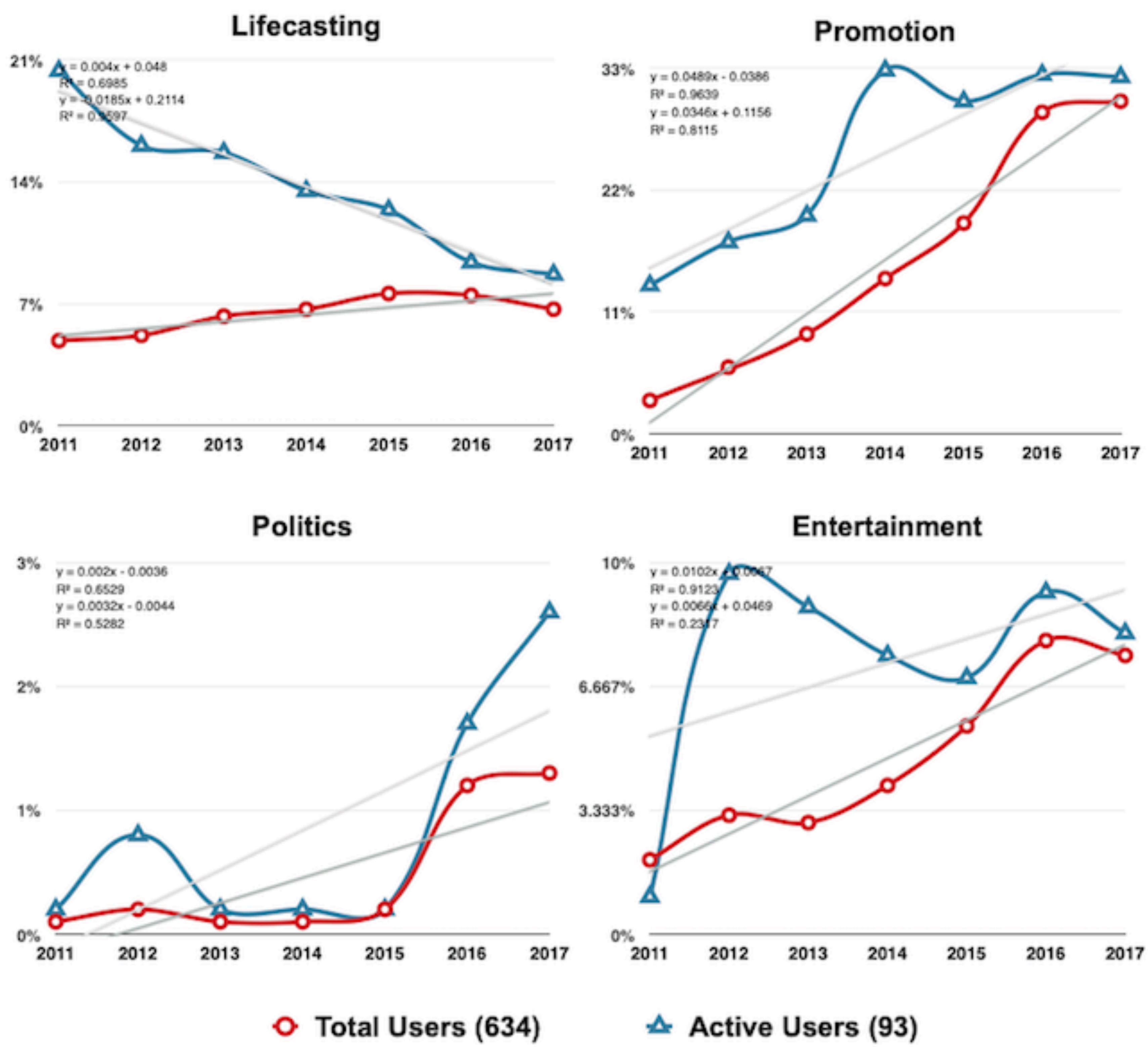

Figure 1: Evolution of weight of four fields.

A repeated-measures ANOVA with a Greenhouse-Geisser correction determined that the weight of the four fields among 637 public Twitter users differed significantly over seven years (*i.e*., Lifecasting: $F(5.14, 3252.89) = 5.30$, $p < .000$; Promotion: $F(3.53, 2234.52) = 65.71$, $p < .000$; Politics: $F(2.10, 1328.68) = 15.18$, $p < .000$; Entertainment: $F(4.09, 2587.09) = 20.08$, $p < .000$). Moreover, the weight of the four fields all demonstrated linearly increasing trends, especially the fields of promotion and entertainment. Post hoc tests using the Bonferroni correction revealed that the weight of lifecasting significantly increased 2 percent from 2011 to 2015, and it slightly decreased after 2015. The weight of promotion significantly increased 10 percent from 2011 to 2014, and then rapidly increased another 15 percent from 2014 to 2016. After 2016, this strong growth momentum declined substantially. The weight of politics peaked in 2017. The weight of entertainment significantly increased 6 percent from 2011 to 2016, and it slightly decreased from 2016 to 2017.

Furthermore, among the 93 Twitter users who tweeted consistently from 2011 to 2017, the weight of three fields differed significantly between seven years (*i.e*., Lifecasting: $F(3.12, 286.85) = 8.45$, $p < .000$; Promotion: $F(4.12, 379.44) = 5.74$, $p < .000$; Politics: $F(1.32, 121.34) = 15.18$, $p = .034$). Lifecasting decreased in salience. The post hoc tests using the Bonferroni correction revealed that the weight of lifecasting significantly dropped 10 percent from 2011 to 2017. The weight of promotion (33 percent) peaked in 2014 and was relatively stable after 2014. Although the weight of politics (2 percent) in 2016 is significantly greater than 2011, 2013, 2014, and 2015, it is not significantly different from the weight in 2012 than in the previous presidential election year.

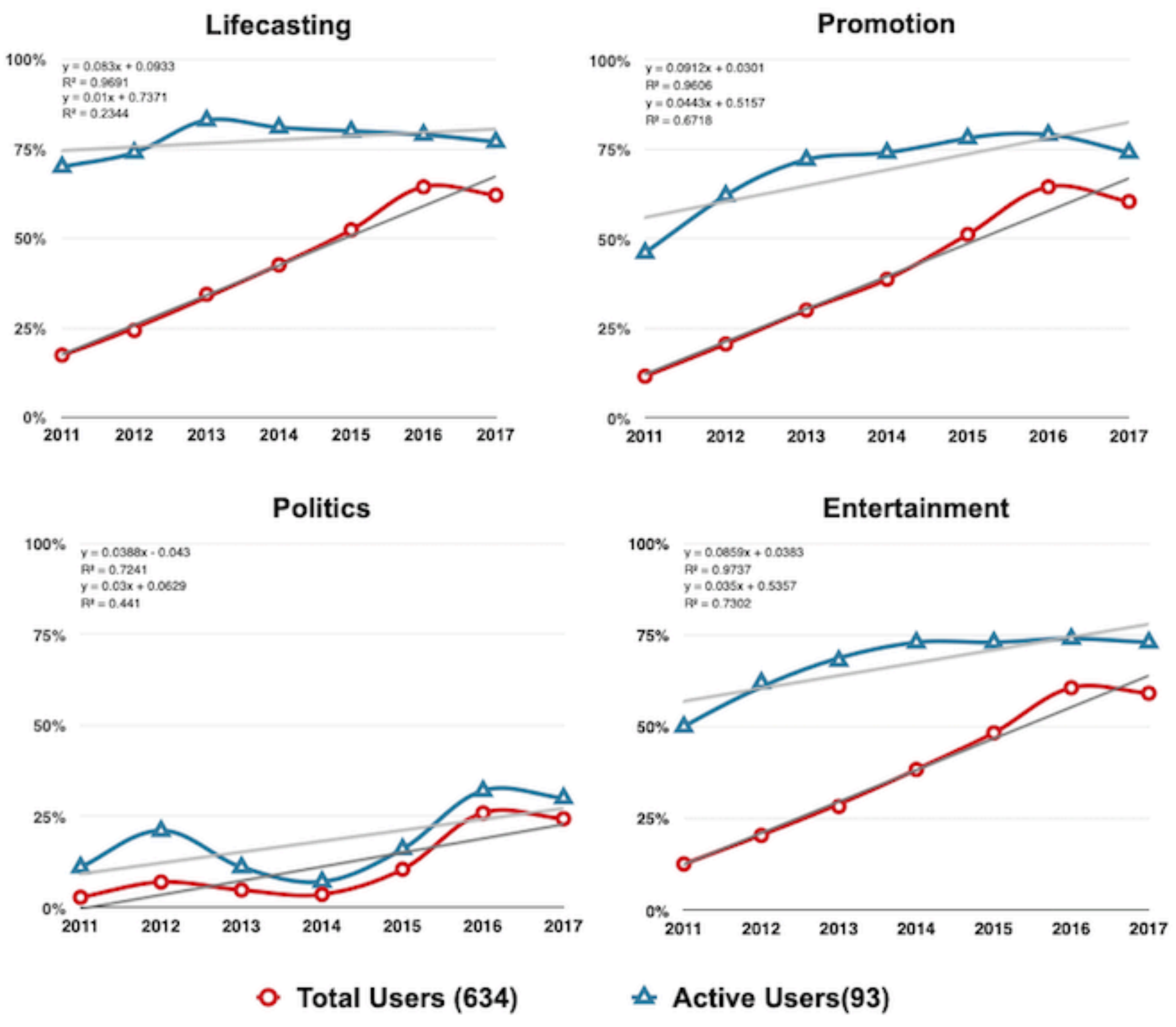

Figure 2: Evolution of user percentage of four fields.

Figure 2 demonstrates the evolution of the ratio of the number of users tweeted on the four fields to the total number of public Twitter users (n = 634), as well as to the number of active users (n = 93) who tweeted in each year. For this, we made four field binary variables per tweet. We assigned value 1 into one of the four binary variables when a tweet has the highest weight over the four fields. Then, we made another four field binary variables per users. When tweet users have at least one binary variable with value 1 per tweet, we dummy-coded the users as value 1. In other words, users could be assigned into multiple fields per year when their tweets have maximum weights from multiple fields.

From 2011 to 2017, among the 637 public Twitter users, the percentage of users who tweeted about lifecasting increased from 17 percent to 62 percent. The percentage of users who tweeted about promotion increased from 11 percent to 60 percent. The percentage of users who tweeted about politics (26 percent) and entertainment (61 percent) all peaked in 2016, indicating the potential influences of 2016 presidential election on the two topics. But among the Twitter users who tweeted consistently from 2011 to 2017 (n = 93), the percentage of users who tweeted about lifecasting was relatively stable; the percentage of users who tweeted about promotion increased about 30 percent; the percentage of users who tweeted on politics and entertainment each increased about 20 percent.

These results indicate that from 2011 to 2017, as more users joined Twitter, they tweeted more in all four fields. However, compared to lifecasting and politics, promotion and entertainment increased more. Although lifecasting is a common field for most Twitter users' tweets, the weight of lifecasting decreased rapidly for the Twitter users that consistently tweeted from 2011 to 2017.

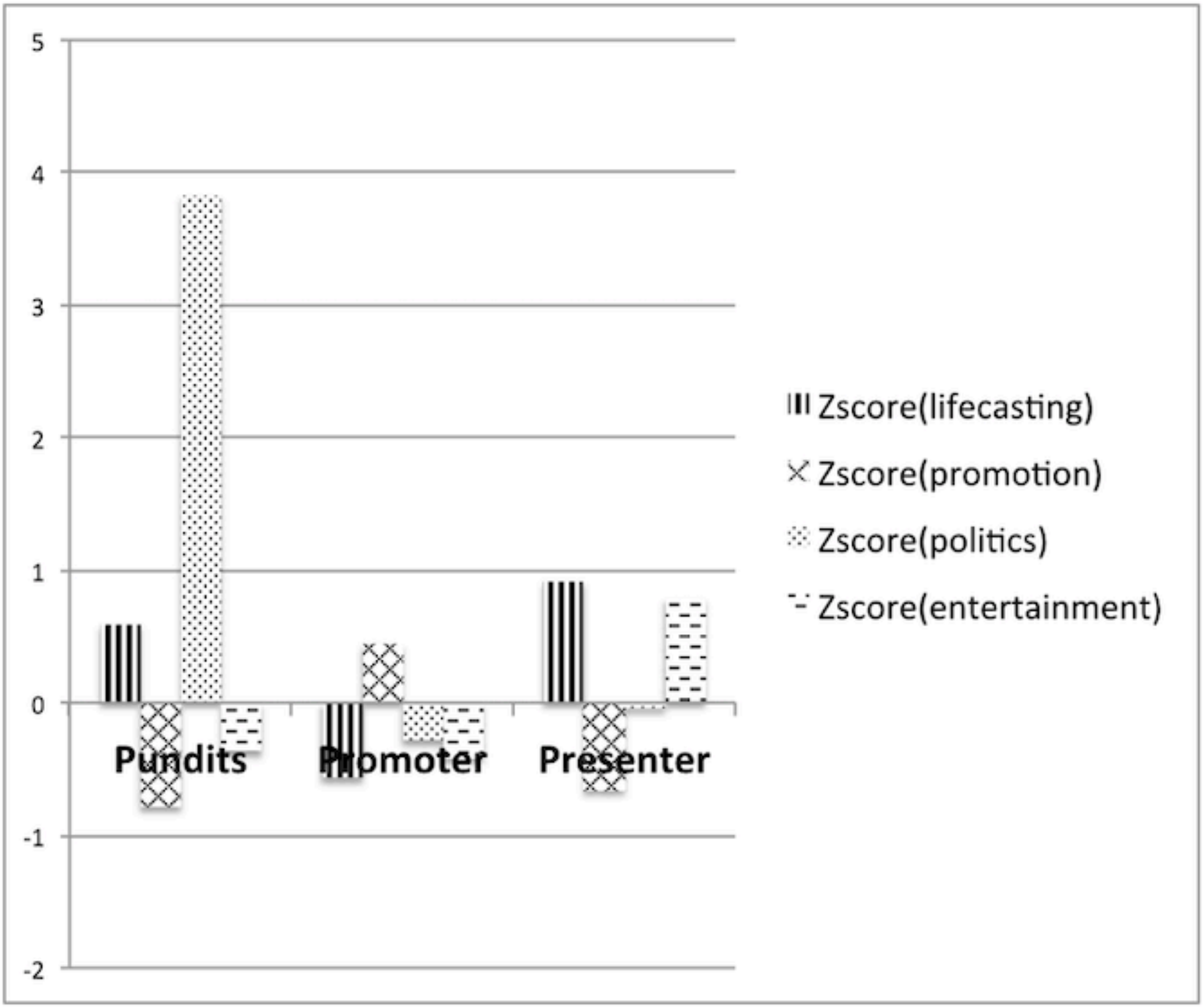

Figure 3: Three types of users participate in four fields.

For *RQ3* (What is the habitus of actors participating in the different online fields of Twitter?), three clusters in the analysis presented the groups most different from each other: Pundits, Promoters, and Presenters (Figure 3). Pundits tweeted more about politics and less about lifecasting. They tweeted less about promotion and entertainment. Promoters only tweeted about promotion. Presenters tweeted about both lifecasting and entertainment, with the value of lifecasting slightly greater than entertainment. Therefore, presenters tweeted about their daily personal life, which also included the entertainment content they were consuming.

Appendix B demonstrates the demographic breakdown of Pundits, Promoters, and Presenters. While more Pundits indicated they were men, most Promoters and Presenters indicated they were women. Latinx and African Americans were less likely than the overall sample to be Pundits, but both groups outpaced the rest of the sample as Presenters. Furthermore, Presenters skewed younger, while Pundits were older. Compared to Pundits and Presenters, Promoters indicated lower levels of education. Nearly half of Pundits reported incomes ranging from US$50,000 to US$74,999. Moreover, the incomes of promoters were lower than Pundits and Presenters.

In terms of sentiment and subjective emotion expressed, we found tweets from Pundits, African Americans, and men were more negative. We also found Presenters expressed significantly more personal emotions and feelings than Promoters. Moreover, education levels were slightly and negatively associated with polarity (r = .01, p = .023), indicating lower-educated people were more positive. Specifically, people with some post-graduate or professional schooling but no postgraduate degree posted the most negative content, while people who have two-year associate degrees were the most positive.

To answer *RQ4* (How is social capital accumulated by different types of actors across different fields?), Appendix C lists the results of comparing production of content and user engagement among three types of social media users and different demographic groups.

We found that the three actors identified for *RQ3* differed significantly in the amount of content they produced. Pundits sent the most direct messages, while Presenters posted the highest number of tweets, retweets, photos, and videos. Promoters, on the other hand, produced the least amount of content. In terms of user engagement, we found that Presenters, a significantly younger group of actors, were more influential than other types of actors as their content was retweeted significantly more, and other Twitter users listed them significantly more often.

We also found African Americans tweeted, retweeted, and sent more messages than Non-African Americans. But their ratio of retweets to tweets (RRP) was higher than Non-African Americans, indicating they produced less original content. In contrast, whites tweeted and retweeted less often than non-whites, but the content they produced was more original. Latinxs and whites produced a similar amount of Twitter content. Latinxs were retweeted significantly more than Non-Latinxs.

Among the four other demographic variables (*e.g.*, age, gender, education, and income), gender was the only significant factor influencing the amount of Twitter content produced. As illustrated in Appendix A, women posted more photos than men. Furthermore, age had weak but significant negative correlations with RRP ($r = .1$, $p < .05$), indicating older Twitter users produced more original content than younger users. Specifically, tweets from users in their 60s were most original, while tweets from users in their 20s were least original. However, the tweets posted by actors in their 20s were retweeted the most.

Modeling

The path model analysis is illustrated in Figure 4. The integrated model fits were acceptable based on the literature (Kline, 2016), showing χ^2 df (10) = 15.796, $p = .106$, root mean square error (RMSEA) = 0.076, comparative fit index (CFI) = 0.956, and standardized root mean square residual (SRMR) = 0.027. The results are illustrated in Figure 4.

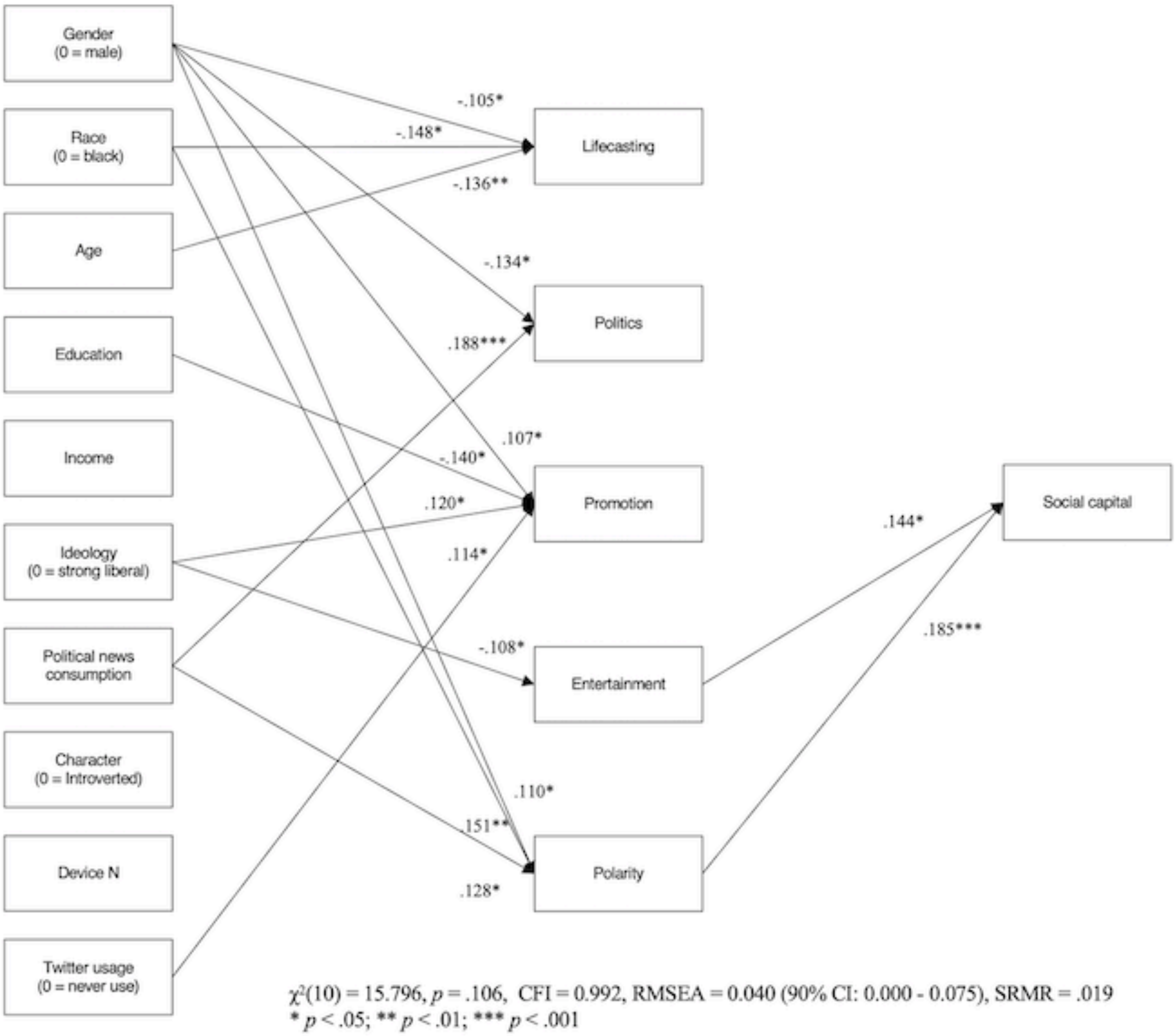

Figure 4: Path model.

We found social capital on Twitter was significantly predicted by entertainment content (β = .144, *p* = .011) and polarity (β = .185, *p* = .001), indicating Twitter users who participated more in the entertainment field, posting positive content, possessed more online social capital. Moreover, tweeting about entertainment was significantly predicted by ideology (β = -.108, *p* = .049), indicating liberal respondents participate more in entertainment. Gender (β = .110, *p* = .40), race (β = .151, *p* = .006), and political news consumption (β = .128, *p* = .020) were positively related to polarity, indicating women, whites, or people with high political news consumption posted more positive tweets.

Besides this, gender (β = -.105, *p* = .047), race (β = -.148, *p* = .006), and age (β = -.136, *p* = .009) significantly predicted participation in the lifecasting field, meaning men, African American, and younger people were more likely to tweet about lifecasting. Political tweets were significantly related to gender (β = -.134, *p* = .012) and political news consumption (β = .188, *p* = .001), with men and those who consumed high levels of political news more likely to participate in the political field. Promotion tweets were statistically significantly related to gender (β = .107, *p* = .041), education (β = -.140, *p* = .012), ideology (β = .120, *p* = .022), and Twitter use frequency (β = .114, *p* = .035). That is, women, those with lower levels of education, conservative ideology, and higher usage of Twitter were more likely to participate in the promotion field.

Discussion and conclusion

In this study we provided a novel research method in social media studies (Stier, *et al*., 2020). While previous studies examining Twitter often use names to infer demographic information like age, gender, and race/ethnicity (Mislove, *et al*., 2011; Sloan, *et al*., 2015; Sloan, *et al*., 2013), here we combined self-identification and self-reported data with individual-level social media data scraped from actual Twitter accounts of survey respondents. The combination of self-reported user profiles and actual posts on Twitter provided more accurate information to define and explore online fields.

Our work extends previous research on the digital production gap by applying Bourdieu's approach to social fields to the second-level divide. We found three types of actors participated in four "interrelated and entangled" fields (Ignatow and Robinson, 2017) with a narrow path toward social capital. As Ignatow and Robinson (2017) note, these fields only make sense in relation to each other. For example, lifecasting, or sharing the daily details about one's personal life, is an important field that spreads effectively on Twitter. Lifecasting tweets are related to the number of times users' tweets are retweeted and favorited by other Twitter users, indicating that tweets in this field are more likely to be shared. However, while younger Twitter users and African Americans employ lifecasting more than older and white users, these tweets are not ultimately related to social capital. Presenters, a significantly younger group of actors, tend to tweet about a combination of lifecasting and entertainment. In fact, the entertainment field is strongly entangled with lifecasting. Nearly 35 percent of the research participants were identified as Presenters who tweeted in both the entertainment and lifecasting fields. Compared to Pundits and Promoters, we found Presenters gain more digital capital by producing more tweets, retweets, photos, and videos. Their tweets were retweeted the most. Furthermore, longitudinally, we found the change of the weight of politics and entertainment, as well as the percentage of Twitter users who tweeted on the two fields are similar, indicating the positive co-evolutions of the two fields. In contrast, we also found that the decrease in participation in the lifecasting field was accompanied by the increase of participation in the promotion field for Twitter users who tweeted consistently in seven years. For these primarily women, older, and more conservative users with less education, Twitter was becoming less personal and more transactional, in line with Robinson's (2018, 2009) findings that users with lower levels of capital tended to use digital platforms more for a task-oriented habitus, while those with higher levels of capital tended to use platforms for more hard play habitus, tweeting more about entertainment.

Promoters do not lifecast. In fact, approximately, 60 percent of our research participants were Promoters who tweeted in the promotion field exclusively. Promotional tweets had significant negative correlations with the number of times users' tweets were retweeted and favorited, indicating that type of content was less influential. Promoters also tweeted much less and are more likely to say in our survey that they "never tweeted." In contrast, the entertainment field, which included tweets about sports and second screening, was the only field significantly related to social capital. Those users with more social capital had the luxury of tweeting about entertainment, while promoters were more concerned with competing in the economic field, sharing transactional and task-oriented "enter to win" types of tweets. These findings have important implications for the future of Twitter. If users with lower skills and education levels use Twitter at increasing levels for transactional purposes as those with more social capital interact with each other in other more play-oriented fields, the digital production gap will continue to widen.

Even in the election year of 2016, politics was not a major field of Twitter for our respondents. In that year, among the four fields, the mean weight of politics is only .04, indicating on average less than 5 percent of the tweets were related to the politics field. Correspondingly, among 579 active Twitter users, only 5 percent (n = 29) were Pundits, whose tweets mainly emphasized politics. While Pundits tended to be older, more men and more middle class, they combined politics with the lifecasting field to increase the influence of their tweets. However, Pundits also tended to send significantly more direct messages, which could indicate that these users were conducting some political discussions in private, which would limit the influence of these discussions from a digital capital standpoint. While a minor field, political engagement on Twitter increased in both weight and numbers of users in 2016. These numbers continued to increase in 2017. This finding seems counterintuitive to the omnipresent political tweets driving American news cycles. However, this field is likely a narrow one, with content dominated by elites and capital not easily obtained.

The findings in this study confirm the conclusion found in existing research that ethnic minorities, especially African Americans, are creating more social media content than U.S. whites (Correa, *et al*., 2010). They retweeted more and sent more messages compared to non-African Americans. However, among 57,996 tweets produced by African Americans, the messages and retweets were significantly more negative than other actors' tweets. The salient negative sentiment emerged from their Twitter conversations may indicate that this group was fighting against repression, echoing a claim from some postmodern researchers (Foucault, 1980; Holtzhausen, 2011, 2000; Kennedy and Sommerfeldt, 2015) that historically oppressed groups refused submission to dominant power by resisting through communication. However, African Americans were not shown to possess higher levels of social capital. If Twitter is to live up to their own stated ideals, improving the platform to incentivize meaningful public deliberation (Jackson and Ibekwe, 2020), they will need to find ways to amplify the voices of the repressed for a more balanced representation of views.

Likewise, we found that the habitus of Pundits and men was significantly more negative. Certainly, the discussion of politics in an increasingly polarized society generated more negative discussions on Twitter. Even so, younger actors and the less educated tended to tweet more positively. Presenters, in particular, expressed more positive personal feelings and emotions than other users. As indicated by how significantly more often this content was shared and favorited, presenters tended to gain more capital using these positive tactics. The entertainment field expresses a more positive outlook than tweets posted in the political field. Perhaps by finding ways to encourage more overlap between the entertainment and political fields, Twitter could begin to narrow the participation gap and at the same time, broaden their online fields.

Limitations and future research

Our research method was not without limitations. We fielded our sample from a paid Qualtrics panel. While we attempted to weight our sample to align with the demographics of Twitter, we were underrepresented among Asian Americans. We were also limited in examining follower networks, which could have provided additional indications of partisanship and interests in different fields. Furthermore, our sample was limited to Twitter users. Future research could expand this same method to include other social media platforms, such as Facebook, Instagram, and Snapchat, to see if these fields, actors, and habitus extend beyond Twitter. Future research should address if the evolution of four fields was different for various demographic groups.

Presenters did not separate the entertainment field from the lifecasting field, which could hold important implications for future research in multitasking and second screening. Likewise, pundits tended to lifecast as well as discuss politics, which could also indicate second screening activity around televised political events such as debates and election coverage.

The prominence of the promotion field in our study indicates that Twitter may have become a significant promotional tool and communication channel for businesses. We confirmed with broader data from Twitter Firehose that this field was not an anomaly in our sample. Future research could address the effectiveness of this field in building digital capital using transactional features of Twitter. This research would investigate Twitter's ability to generate revenue, a field that did not prove influential in our sample and an area in which Twitter has faced challenges.

About the authors

Jiang Ke is assistant professor of communication design at Elon University based in Elon, North Carolina.
E-mail: jjiang [at] elon [dot] edu

Lance Porter is the is the Class of 1941 Alumni Professor and the founding director of the Social Media Analysis and Creation (SMAC) Lab in the Manship School of Mass Communication at Louisiana State University.
Send comments to: lporter [at] lsu [dot] edu

Rui Wang is an associate professor in the Institute of Communication Studies at the Communication University of China.
E-mail: ruiwang [dot] communication [at] gmail [dot] com

Seon-Woo Kim is a Ph.D. student in the Manship School of Mass Communication and a M.S. student of analytics at Louisiana State University.
E-mail: kseonw1 [at] lsu [dot] edu

Martin Johnson was the Kevin P. Reilly, Sr. Chair in Political Communication and Professor of Mass Communication and Political Science in the Manship School of Mass Communication at Louisiana State

University.

Notes

1. van Dijk, 2006, p. 221.

2. Helsper and Reisdorf, 2017, p. 1,265.

3. Lindell, 2017, p. 1.

4. Ignatow and Robinson, 2017, p. 952.

References

A. Amelia, 2018. "K-means clustering: From a to z everything you need to know about k-means clustering" (27 September), at https://towardsdatascience.com/k-means-clustering-from-a-to-z-f6242a314e9a, accessed 18 November 2022.

Y. Arie and G.S. Mesch, 2015. "Interethnic ties via mobile communications in homogeneous and ethnically mixed cities: A structural diversification approach," *Communication and Information Technologies Annual*, volume 10, pp. 175–203.
doi: https://doi.org/10.1108/S2050-206020150000010007, accessed 18 November 2022.

Y. Benkler and H. Nissenbaum, 2006. "Commons-based peer production and virtue," *Journal of Political Philosophy*, volume 14, number 4, pp. 394–419.
doi: https://doi.org/10.1111/j.1467-9760.2006.00235.x, accessed 18 November 2022.

D.M. Blei, A.Y. Ng, and M.I. Jordan, 2003. "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, volume 3, pp. 993–1,022, and at https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf, accessed 18 November 2022.

P. Bourdieu, 2004. *Science of science and reflexivity*. Translated by R. Nice. Chicago: University of Chicago Press.

P. Bourdieu, 1990. *The logic of practice*. Translated by R. Nice. Stanford, Calif.: Stanford University Press.

P. Bourdieu, 1986. "The forms of capital," In: J.G. Richardson (editor). *Handbook of theory and research for the sociology of education*. Westport, Conn.: Greenwood Press, pp. 241–258.

P. Bourdieu, 1984. *Distinction: A social critique of the judgement of taste*. Translated by R. Nice. Cambridge, Mass.: Harvard University Press.

D.R. Brake, 2014. "Are we all online content creators now? Web 2.0 and digital divides," *Journal of Computer-Mediated Communication*, volume 19, number 3, pp. 591–609.
doi: https://doi.org/10.1111/jcc4.12042, accessed 18 November 2022.

A. Bruns, 2008. *Blogs, Wikipedia, Second Life, and beyond: From production to produsage*. New York: Peter Lang.

T. Correa, A.W. Hinsley, and H.G. de Zúñiga, 2010. "Who interacts on the Web? The intersection of users' personality and social media use," *Computers in Human Behavior*, volume 26, number 2, pp. 247–253.
doi: https://doi.org/10.1016/j.chb.2009.09.003, accessed 18 November 2022.

J. Diesner, A. Aleyasen, J. Kim, S. Mishra, and K. Soltani, 2013. "Using socio-semantic network analysis for assessing the impact of documentaries," paper presented at the Workshop of Information in Networks at https://jdiesnerlab.ischool.illinois.edu/publications/WIN_2013_09132013_final.pdf, accessed 18 November 2022.

M. Foucault, 1980. *Power/knowledge: Selected interviews and other writings 1972–1977*. Translated by C.

Gordon, L. Marshall, J. Mepham, and K. Sopers. New York: Vintage.

M. Gilbert, 2010. "Theorizing digital and urban inequalities: Critical geographies of 'race', gender and technological capital," *Information, Communication & Society*, volume 13, number 7, pp. 1,000–1,018.
doi: https://doi.org/10.1080/1369118X.2010.499954, accessed 18 November 2022.

S. Halford and M. Savage, 2010. "Reconceptualizing digital social inequality," *Information, Communication & Society*, volume 13, number 7, pp. 937–955.
doi: https://doi.org/10.1080/1369118X.2010.499956, accessed 18 November 2022.

E. Hargittai and A. Hinnant, 2008. "Digital inequality: Differences in young adults' use of the Internet," *Communication Research*, volume 35, number 5, pp. 602–621.
doi: https://doi.org/10.1177/0093650208321782, accessed 18 November 2022.

C. Harris, L. Straker, and C. Pollock, 2017. "A socioeconomic related 'digital divide' exists in how, not if, young people use computers," *PLoS ONE*, volume 12, number 3, e0175011.
doi: https://doi.org/10.1371/journal.pone.0175011, accessed 18 November 2022.

E.J. Helsper and B.C. Reisdorf, 2017. "The emergence of a 'digital underclass' in Great Britain and Sweden: Changing reasons for digital exclusion," *New Media & Society*, volume 19, number 8, pp. 1,253–1,270.
doi: https://doi.org/10.1177/1461444816634676, accessed 18 November 2022.

M. Hofer and V. Aubert, 2013. "Perceived bridging and bonding social capital on Twitter: Differentiating between followers and followees," *Computers in Human Behavior*, volume 29, number 6, pp. 2,134–2,142.
doi: https://doi.org/10.1016/j.chb.2013.04.038, accessed 18 November 2022.

C.P. Hoffmann, C. Lutz, and M. Meckel, 2015. "Content creation on the Internet: A social cognitive perspective on the participation divide," *Information, Communication & Society*, volume 18, number 6, pp. 696–716.
doi: https://doi.org/10.1080/1369118X.2014.991343, accessed 18 November 2022.

D.R. Holtzhausen, 2011. *Public relations as activism: Postmodern approaches to theory and practice*. New York: Routledge.
doi: https://doi.org/10.4324/9780203819098, accessed 18 November 2022.

D.R. Holtzhausen, 2000. "Postmodern values in public relations," *Journal of Public Relations Research*, volume 12, number 1, pp. 93–114.
doi: https://doi.org/10.1207/S1532754XJPRR1201_6, accessed 18 November 2022.

G. Ignatow and L. Robinson, 2017. "Pierre Bourdieu: Theorizing the digital," *Information, Communication & Society*, volume 20, number 7, pp. 950–966.
doi: https://doi.org/10.1080/1369118X.2017.1301519, accessed 18 November 2022.

L. Jackson with D. Ibekwe, 2020. Jack Dorsey on Twitter's mistakes: The social network's C.E.O. has plans to improve the platform. But will the changes be enough?" *New York Times* (7 August), at https://www.nytimes.com/2020/08/07/podcasts/the-daily/Jack-dorsey-twitter-trump.html, accessed 18 November 2022.

S. Jain and A. Sinha, 2020. "Identification of influential users on Twitter: A novel weighted correlated influence measure for Covid-19," *Chaos, Solitons & Fractals*, volume 139, 110037.
doi: http://doi.org/10.1016/j.chaos.2020.110037, accessed 18 November 2022.

H. Jenkins, 2006. *Convergence culture: Where old and new media collide*. New York: New York University Press.

H. Jenkins, with K. Clinton, R. Purushotma, A.J. Robison, and M. Weigel, 2009. *Confronting the challenges of participatory culture: Media education for the 21st century*. Cambridge, Mass.: MIT Press.
doi: https://doi.org/10.7551/mitpress/8435.001.0001, accessed 18 November 2022.

A.K. Kennedy and E.J. Sommerfeldt, 2015. "A postmodern turn for social media research: Theory and research directions for public relations scholarship," *Atlantic Journal of Communication*, volume 23, number 1, pp. 31–45.
doi: https://doi.org/10.1080/15456870.2015.972406, accessed 18 November 2022.

V.A. Kharde and S.S. Sonawane, 2016. "Sentiment analysis of Twitter data: A survey of techniques," *International Journal of Computer Applications*, volume 139, number 11, pp. 5–11.
doi: https://doi.org/10.5120/ijca2016908625, accessed 18 November 2022.

R.B. Kline, 2016. *Principles and practice of structural equation modeling* Fourth edition. New York: Guilford Press.

C. Leadbeater, 2008. *We-think: Mass innovation, not mass production*. London: Profile.

N. Levina and M. Arriaga, 2014. "Distinction and status production on user-generated content platforms: Using Bourdieu's theory of cultural production to understand social dynamics in online fields," *Information Systems Research*, volume 25, number 3, pp. 468–488.
doi: https://doi.org/10.1287/isre.2014.0535, accessed 18 November 2022.

K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, 2008. "Tastes, ties, and time: A new social network dataset using Facebook.com," *Social Networks*, volume 30, number 4, pp. 330–342.
doi: https://doi.org/10.1016/j.socnet.2008.07.002, accessed 18 November 2022.

J. Lindell, 2017. "Bringing field theory to social media, and vice-versa: Network-crawling an economy of recognition on Facebook," *Social Media + Society*, volume 3, number 4 (17 October).
doi: https://doi.org/10.1177/2056305117735752, accessed 18 November 2022.

C. Lutz, 2016. "A social milieu approach to the online participation divides in Germany," *Social Media + Society*, volume 2, number 1 (21 January).
doi: https://doi.org/10.1177/2056305115626749, accessed 18 November 2022.

M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik, 2021. *Cluster: Cluster analysis basics and extensions*, R package, version 2.1.2.

R. McGrath, 2013. "Twython 3.8.0 documentation," at
https://twython.readthedocs.io/en/latest/usage/basic_usage.html, accessed 18 November 2022.

G.S. Mesch and I. Talmud, 2011. "Ethnic differences in Internet access: The role of occupation and exposure," *Information, Communication & Society*, volume 14, number 4, pp. 445–471.
doi: https://doi.org/10.1080/1369118X.2011.562218, accessed 18 November 2022.

S. Mihelj, A. Leguina, and J. Downey, 2019. "Culture is digital: Cultural participation, diversity and the digital divide," *New Media & Society*, volume 21, number 7, pp. 1,465–1,485.
doi: https://doi.org/10.1177/1461444818822816, accessed 18 November 2022.

A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, and N. Rosenquist, 2011. "Understanding the demographics of Twitter users," *Proceedings of the International AAAI Conference on Weblogs and Social Media*, volume 5, number 1, pp. 554–557.
doi: https://doi.org/10.1609/icwsm.v5i1.14168, accessed 18 November 2022.

B. O'Connor, R. Balasubramanyan, B.R. Routledge, and N.A. Smith, 2010. "From tweets to polls: Linking text sentiment to public opinion time series," *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, at https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1536/1842, accessed 18 November 2022.

Meltem Odabaş, 2022. "10 facts about Americans and Twitter," *Pew Research Center* (5 May), at https://www.pewresearch.org/fact-tank/2022/05/05/10-facts-about-americans-and-twitter/, accessed 18 November 2022.

H. Ono and M. Zavodny, 2008. "Immigrants, English ability and the digital divide," *Social Forces*, volume 86, number 4, pp. 1,455–1,479.
doi: https://doi.org/10.1353/sof.0.0052, accessed 18 November 2022.

Pew Research Center, 2021. "Social media fact sheet" (7 April), at https://www.pewresearch.org/internet/fact-sheet/social-media/, accessed 18 November 2022.

J. Phua, S.V. Jin, and J. Kim, 2017. "Uses and gratifications of social networking sites for bridging and bonding social capital: A comparison of Facebook, Twitter, Instagram, and Snapchat," *Computers in Human Behavior*, volume 72, pp. 115–122.
doi: https://doi.org/10.1016/j.chb.2017.02.041, accessed 18 November 2022.

B.C. Reisdorf and D. Groselj, 2017. "Internet (non-)use types and motivational access: Implications for digital inequalities research," *New Media & Society*, volume 19, number 8, pp. 1,157–1,176.
doi: https://doi.org/10.1177/1461444815621539, accessed 18 November 2022.

M.V. Reiss and M. Tsvetkova, 2020. "Perceiving education from Facebook profile pictures," *New Media & Society*, volume 22, number 3, pp. 550–570.
doi: https://doi.org/10.1177/1461444819868678, accessed 18 November 2022.

F.J. Riggins and S. Dewan, 2005. "The digital divide: Current and future research directions," *Journal of the Association for Information Systems*, volume 6, number 12.
doi: https://doi.org/10.17705/1jais.00074, accessed 18 November 2022.

G. Ritzer and N. Jurgenson, 2010. "Production, consumption, prosumption: The nature of capitalism in the age of the digital 'prosumer'," *Journal of Consumer Culture*, volume 10, number 1, pp. 13–36.
doi: https://doi.org/10.1177/1469540509354673, accessed 18 November 2022.

L. Robinson, 2018. "The identity curation game: Digital inequality, identity work, and emotion management," *Information, Communication & Society*, volume 21, number 5, pp. 661–680.
doi: https://doi.org/10.1080/1369118X.2017.1411521, accessed 18 November 2022.

L. Robinson, 2009. "A taste for the necessary: A Bourdieuian approach to digital inequality," *Information, Communication & Society*, volume 12, number 4, pp. 488–507.
doi: https://doi.org/10.1080/13691180902857678, accessed 18 November 2022.

L. Robinson, S.R. Cotten, H. Ono, A. Quan-Haase, G. Mesch, W. Chen, J. Schulz, T.M. Hale, and M.J. Stern, 2015. "Digital inequalities and why they matter," *Information, Communication & Society*, volume 18, number 5, pp. 569–582.
doi: https://doi.org/10.1080/1369118X.2015.1012532, accessed 18 November 2022.

J.M. Robles Morales, M. Antino, S. De Marco, and J.A. Lobera, 2016. "La nueva frontera de la desigualdad digital: la brecha participativa (The new frontier of digital inequality: The participatory divide)," *Revista Española de Investigaciones Sociologicas*, número 156, pp. 97–115.
doi: https://doi.org/10.5477/cis/reis.156.97, accessed 18 November 2022.

Y. Rosseel, 2012. "lavaan: An R package for structural equation modeling," *Journal of Statistical Software*, volume 48, number 2, pp. 1–36.
doi: https://doi.org/10.18637/jss.v048.i02, accessed 18 November 2022.

J. Schradie, 2011. "The digital production gap: The digital divide and Web 2.0 collide," *Poetics*, volume 39,number 2, pp. 145–168.
doi: https://doi.org/10.1016/j.poetic.2011.02.003, accessed 18 November 2022.

P. Shah, 2020. "Sentiment analysis using textblob" (27 June), at https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524, accessed 18 November 2022.

C. Sims, 2014. "From differentiated use to differentiating practices: Negotiating legitimate participation and the production of privileged identities," *Information, Communication & Society*, volume 17, number 6, pp. 670–682.
doi: https://doi.org/10.1080/1369118X.2013.808363, accessed 18 November 2022.

L. Sloan, J. Morgan, P. Burnap, and M. Williams, 2015. "Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data," *PLoS ONE*, volume 10, number 3, e0115545.
doi: https://doi.org/10.1371/journal.pone.0115545, accessed 18 November 2022.

L. Sloan, J. Morgan, W. Housley, M. Williams, A. Edwards, P. Burnap, and O. Rana, 2013. "Knowing the tweeters: Deriving sociologically relevant demographics from Twitter," *Sociological Research Online*, volume 18, number 3,

pp. 74–84.
doi: https://doi.org/10.5153/sro.3001, accessed 18 November 2022.

S. Stier, J. Breuer, P. Siegers, and K. Thorson, 2020. “Integrating survey data and digital trace data: Key issues in developing an emerging field,” *Social Science Computer Review*, volume 38, number 5, pp. 503–516.
doi: https://doi.org/10.1177/0894439319843669, accessed 18 November 2022.

L. Townsend and C. Wallace, 2016. “Social media research: A guide to ethics,” at https://www.bolton.ac.uk/assets/Uploads/Social-media-ethics-study-Aberdeen-2018.pdf, accessed 18 November 2022.

J.A.G.M. van Dijk, 2006. “Digital divide research, achievements and shortcomings,” *Poetics*, volume 34, numbers 4–5, pp. 221–235.
doi: https://doi.org/10.1016/j.poetic.2006.05.004, accessed 18 November 2022.

M.L. Williams, P. Burnap, and L. Sloan, 2017. “Towards an ethical framework for publishing Twitter data in social research: Taking into account users’ views, online context and algorithm estimation,” *Sociology*, volume 51, number 6, pp. 1,149–1,168.
doi: https://doi.org/10.1177/0038038517708140, accessed 18 November 2022.

J.C. Witte and S.E. Mannon, 2010. *The Internet and social inequalities*. New York: Routledge.
doi: https://doi.org/10.4324/9780203861639, accessed 18 November 2022.

S. Wojcik and A. Hughes, 2019. “Sizing up Twitter users,” *Pew Research Center* (24 April), at https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/, accessed 18 November 2022.

U. Yaqub, N. Sharma, R. Pabreja, S.A. Chun, V. Atluri, and J. Vaidya, 2018. “Analysis and visualization of subjectivity and polarity of Twitter location data,” *dg.o ’18: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, article number 67, pp. 1–10.
doi: https://doi.org/10.1145/3209281.3209313, accessed 18 November 2022.

S. Yates and E. Lockley, 2018. “Social media and social class,” *American Behavioral Scientist*, volume 62, number 9, pp. 1,291–1,316.
doi: https://doi.org/10.1177/0002764218773821, accessed 18 November 2022.

N. Zillien and M. Marr, 2013. “The digital divide in Europe,” In: M. Ragnedda and G. Muschert (editors). *The digital divide: The Internet and social inequality in international perspective*. London: Routledge, pp. 55–66.
doi: https://doi.org/10.4324/9780203069769, accessed 18 November 2022.

Appendix A: Production of content around four fields by demographic groups.
Note: LC: Lifecasting; PR: Promotion; PO: Politics; EN: Entertainment
* *p* <.05, ** *p* <.01, *** *p* <.001

			LC	PR*	PO**	EN
Gender	**Men**	M	0.277	0.339	**0.057**	0.117
	(n = 186)	*SD*	*0.224*	*0.347*	*0.138*	*0.199*
	Women	M	0.258	**0.413**	0.028	0.111
	(n = 393)	*SD*	*0.236*	*0.361*	*0.078*	*0.186*
			LC*	PR	PO	EN
	Black	M	**0.305**	0.346	0.028	0.106
	(n =					

Race1	186)	*SD*	*0.239*	*0.340*	*0.074*	*0.192*
	NBlack	M	0.249	0.405	0.041	0.115
	(n= 393)	*SD*	*0.229*	*0.363*	*0.110*	*0.190*
Race2			LC*	PR	PO	EN
	White	M	0.247	0.406	0.040	0.113
	(n = 394)	*SD*	*0.229*	*0.361*	*0.108*	*0.187*
	NWhite	M	**0.300**	0.352	0.033	0.113
	(n= 185)	*SD*	*0.236*	*0.349*	*0.087*	*0.198*
Ethnicity			LC	PR	PO	EN
	Hisp	M	0.271	0.344	0.026	0.120
	(n = 156)	*SD*	*0.246*	*0.361*	*0.082*	*0.203*
	NHisp	M	0.261	0.406	0.041	0.110
	(n= 423)	*SD*	*0.227*	*0.355*	*0.108*	*0.186*
Age			LC***	PR***	PO*	EN
	20s	M	**0.348**	0.262	0.035	0.104
	(n = 130)	*SD*	*0.252*	*0.308*	*0.078*	*0.172*
	30s	M	0.259	0.421	0.025	0.121
	(n= 209)	*SD*	*0.240*	*0.357*	*0.073*	*0.200*
	40s	M	0.225	0.422	0.036	0.134
	(n = 118)	*SD*	*0.205*	*0.360*	*0.102*	*0.222*
	50s	M	0.205	**0.491**	0.050	0.085
	(n= 75)	*SD*	*0.189*	*0.379*	*0.142*	*0.149*
	60s	M	0.242	0.356	**0.084**	0.091
	(n= 44)	*SD*	*0.212*	*0.374*	*0.169*	*0.170*
			LC	PR**	PO	EN
	≤High school	M	0.256	**0.466**	0.033	0.083
	(n = 125)	*SD*	*0.243*	*0.365*	*0.100*	*0.164*
	Some college	M	0.264	0.408	0.033	0.114
	(n = 161)	*SD*	*0.223*	*0.362*	*0.085*	*0.199*
	Two year	M	0.245	0.431	0.036	0.129

Education	*(n = 91)*	*SD*	*0.227*	*0.356*	*0.107*	*0.204*
	BS, BA, AB	M	0.283	0.327	0.041	0.130
	(n = 136)	*SD*	*0.233*	*0.343*	*0.097*	*0.205*
	Some post	M	0.159	0.332	0.011	0.098
	(n = 15)	*SD*	*0.199*	*0.380*	*0.036*	*0.179*
	≥MA, MS	M	0.291	0.250	0.062	0.107
	(n = 51)	*SD*	*0.251*	*0.303*	*0.155*	*0.158*
Income			LC	PR	PO*	EN
	Mid class	M	0.272	0.357	**0.055**	0.127
	(n = 134)	*SD*	*0.230*	*0.357*	*0.129*	*0.212*
	Others	M	0.261	0.399	0.032	0.108
	(n= 445)	*SD*	*0.233*	*0.358*	*0.091*	*0.183*

Appendix B: Demographics of three user types.

Demographics		**Pundits** (N = 29)	**Promoters** (N = 350)	**Presenters** (N = 200)
Gender	Men	55.2%	28.9%	34.5%
	Women	44.8%	71.1%	65.5%
Ethnicity	Hispanic, Latinx or Spanish origin	13.8%	26%	30.5%
	Non-Hispanic	86.2%	74%	69.5%
Race	Black or African American	20.7%	25.7%	29%
	Nonblack	79.3%	74.3%	71%
	White	69%	69.7%	65%
	Nonwhite	31%	30.3%	35%
	20–29	20.7%	16.9%	32.5%
	30–39	20.7%	38%	35%

Age	40–49	20.7%	21.7%	18%
	50–59	17.2%	15.1%	8.5%
	60 and older	20.7%	7.7%	5.5%
Education	High school or less	20.7%	24.6%	16.5%
	Some college	27.6%	26.6%	30%
	Two year associate degree	13.8%	16.3%	15%
	BS, BA, AB	27.6%	21.7%	26%
	Some postgraduate but no degree	0%	3.4%	1.5%
	MA, MS, PHD, MD, JD	10.3%	7.4%	11%
Income	Less than US$14,999	3.4%	12.3%	7.5%
	US$15,000 to US$24,999	6.9%	10%	10%
	US$25,000 to US$34,999	20.7%	13.4%	16%
	US$35,000 to US$49,999	10.3%	18%	18%
	US$50,000 to US$74,999	41.4%	21.1%	24%
	US$75,000 to US$99,999	6.9%	10.9%	12.5%
	US$100,000 to US$149,000	10.3%	8.6%	8.5%
	US$150,000 to US$199,999	0%	2.3%	1.5%
	US$200,000 or more	0%	1.7%	0.5%

Appendix C: Capital and sentiments by demographic group and type of social media actor.
Note: T: Tweets; RT: Retweet; M: Message; P: Photo; V: Video; FER: Follower; LID: LID: Listed by other users as a

member of a social group; RTD: Retweeted; FED: Favorited; SE: polarity; EM: Subjectivity
* p <.05, ** p <.01, *** p <.001

			Capital: Amount of Twitter content						Capital: User engagement				Sentiments	
Gender			T	RT	M	P**	V	RRP	FER	LID	RTD	FED	SE*	EM
	Men	M	242	56	53	32	9	0.12	376	12	395	1	0.16	0.33
	(n = 186)	SD	450	165	219	87	46	0.21	1682	70	1342	4	0.17	0.18
	Women	M	280	76	31	**56**	5	0.16	1174	15	824	1	**0.20**	0.34
	(n= 393)	SD	475	229	100	111	37	0.33	13772	82	4663	9	0.19	0.17
Race 1			T**	RT*	M*	P	V	RRP*	FER	LID	RTD	FED	SE**	EM
	Black	M	**362**	**106**	**59**	45	9	**0.2**	750	15	543	0	0.15	0.32
	(n = 186)	SD	591	288	229	91	49	0.44	4706	74	2078	1	0.17	0.18
	NBlack	M	234	56	30	49	6	0.13	978	14	738	1	**0.20**	0.34
	(n= 393))	SD	409	173	106	109	37	0.23	12992	80	4401	9	0.19	0.17
Race 2			T*	RT*	M	P	V	RRP*	FER	LID	RTD	FED	SE	EM
	White	M	235	57	32	48	6	0.13	1037	14	536	1	0.20	0.34
	(n = 394)	SD	415	173	110	106	38	0.22	13491	83	1796	9	0.18	0.17
	NWhite	M	**339**	**97**	51	49	9	**0.19**	665	13	1006	0	0.17	0.33
	(n= 185)	SD	557	273	210	100	44	0.42	4304	67	6420	1	0.19	0.18
Ethnicity			T	RT	M	P	V	RRP	FER	LID	RTD*	FED*	SE	EM
	Hisp	M	246	78	45	53	11	0.16	2347	24	**1297**	**2**	0.19	0.32
	(n = 156)	SD	470	253	219	114	59	0.26	21406	126	7027	14	0.19	0.17
	NHisp	M	276	66	36	47	5	0.15	391	10	461	0	0.19	0.34
	(n= 423)	SD	466	193	113	101	30	0.31	2879	50	1647	1	0.18	0.18
Age			T	RT	M	P	V	RRP**	FER	LID	RTD**	FED	SE	EM
	20s	M	290	106	48	44	6	**0.23**	774	15	**2017**	0	0.16	0.34
	(n = 130)	SD	527	237	224	87	16	0.30	5129	84	7898	1	0.17	0.18
	30s	M	256	48	44	45	10	0.12	446	13	404	1	0.20	0.34
	(n= 209)	SD	418	135	144	94	58	0.21	1644	47	1391	4	0.20	0.17
	40s	M	285	68	34	70	3	0.16	2527	19	186	2	0.19	0.35
	(n = 118)	SD	477	216	99	141	9	0.46	24520	130	660	16	0.19	0.18
	50s	M	309	95	30	46	8	0.12	458	16	182	0	0.18	0.32
	(n= 75)	SD	528	338	111	87	50	0.22	1845	46	552	0	0.17	0.18
	60s	M	164	25	7	27	1	0.09	116	4	343	1	0.21	0.32
	(n= 44)	SD	365	77	13	113	2	0.17	216	9	1126	2	0.21	0.19
			T	RT	M	P	V	RRP	FER***	LID***	RTD	FED***	SE**	EM
	≤High School	M	259	47	25	45	6	0.12	243	4	1013	0	0.22	0.35
	(n = 125)	SD	3429	150	67	96	43	0.23	1369	9	7409	0	0.17	0.17
	Some College	M	301	84	54	52	7	0.18	350	13	614	0	0.17	0.32

Education	*(n = 161)*	SD	3510	235	236	105	46	0.42	952	46	2131	1	0.15	0.18
	Two Year	M	243	54	27	51	8	0.16	299	13	510	0	**0.23**	0.35
	(n = 91)	SD	3391	130	84	112	47	0.25	1096	89	1881	1	0.20	0.17
	BS, BA, AB	M	267	83	42	51	7	0.16	495	15	667	0	0.18	0.35
	(n = 136)	SD	3494	262	115	110	34	0.25	1737	51	2435	1	0.21	0.17
	Some Post	M	212	22	7	34	9	0.08	**17939**	**112**	448	**12**	0.05	0.25
	(n = 15)	SD	3436	44	15	83	26	0.23	68728	363	1666	44	0.26	0.26
	≥MA, MS	M	250	84	38	38	2	0.12	1590	12	550	1	0.16	0.33
	(n = 51)	SD	486	245	134	101	5	0.20	8243	44	1836	8	0.19	0.17
Income			T	RT	M**	P	V	RRP	FER	LID	RTD	FED	SE**	EM
	Mid Class	M	251	70	27	48	6	0.15	516	12	792	0	0.19	0.34
	(n = 134)	SD	451	221	95	105	37	0.32	3112	55	4435	3	0.18	0.18
	Others	M	325	68	**75**	51	8	0.14	2253	21	335	2	0.18	0.33
	(n = 445)	SD	513	173	255	103	50	0.21	22998	128	972	15	0.19	0.16
Types of social media users			T***	RT***	M**	P**	V*	RRP***	FER	LID*	RTD**	FED	SE**	EM**
	Pundits	M	286	56	90	22	3	0.26	98	4	924	1	0.11	0.35
	(n = 29)	SD	456	109	212	64	5	0.27	154	9	1694	1	0.13	0.17
	Presenters	M	380	116	62	69	13	0.26	1931	28	1521	2	0.16	0.37
	(n = 200)	SD	556	281	210	134	58	0.42	18878	126	6478	13	0.13	0.12
	Promoters	M	203	44	20	39	3	0.08	407	7	189	0	0.21	0.32
	(n = 350)	SD	395	160	85	85	26	0.17	3266	29	796	1	0.21	0.20

Appendix D: List of data variables used.

Online fields		**Survey data**
LC **PR** **PO** **EN**	Lifecasting Promotion Politics Entertainment	Gender Ethnicity Race Age Education Income Political news consumption Political ideology
Twitter content production		
T **RT** **M** **P** **V**	Tweet Retweet Messages Photo Video	
Twitter engagement		
FER **LID** **RED**	Follower Listed by other users as a member of a social group Retweeted	

FED	Favorited	
Twitter sentiment and emotion		
SE **EM**	Polarity Subjectivity	

Editorial history

Received 12 July 2021; revised 27 July 2022; accepted 8 November 2022.

Pundits, presenters, and promoters: Investigating gaps in digital production among social media users using self-reported and behavioral measures
by Jiang Ke, Lance Porter, Rui Wang, Seon-Woo Kim, and Martin Johnson.
First Monday, Volume 27, Number 12 - 5 December 2022
https://firstmonday.org/ojs/index.php/fm/article/download/11604/10748
doi: https://dx.doi.org/10.5210/fm.v27i12.11604