Examining the Social Media Antecedents of Racial Justice:
Evidence from Twitter

Zhasmina Tacheva
Syracuse University
ztacheva@syr.edu

Leni Krsova
Syracuse University
lkrsova@syr.edu

Anton Ivanov
University of Illinois at Urbana-Champaign
antoniva@illinois.edu

Abstract

Despite an increasing body of research focusing on the use of social media posts for inferring various social phenomena such as influenza spread, political preferences, or information diffusion, less attention has been given to the influence such posts have on user attitudes. Rooted at the nexus of social capital, sociolinguistics, and the theory of critical consciousness, this study aims to explore the social media factors playing a role in racial justice attitude development. Using a matched sample of 5,365 Twitter users, we demonstrate the impact of user posts' topical diversity and bridging and bonding capital on the likelihood of engaging in antiracist discourse online in the aftermath of George Floyd's murder.

1. Introduction

Captured in a shocking video gone viral, the last moments of 46-year-old African American husband and father George Floyd in Minneapolis, MN, on the night of May 25, 2020, sparked nationwide protests and demonstrations in 80 countries outside the United Sates. The struggles for justice of Black, Indigenous, and people of color (BIPOC) were met with an unprecedented number of allies willing to listen and educate themselves on issues of race, mourn the losses of BIPOC communities, critically interrogate their own privilege, support BIPOC-owned businesses, and engage in antiracist practices—a phenomenon collectively referred to in Black American parlance as "becoming woke." In this context, in spite of well-founded criticism of optical allyship and slacktivism, social media has emerged as a vital conduit of information, allowing users not just to manifest their own thoughts and feelings but also to learn from others.

Since Hurricane Sandy in 2012, social media channels have been used extensively to disseminate information during critical events due to their instantaneousness and resilience compared to traditional media [1]. Social media is particularly useful in connecting volunteers for collective action, both when it comes to natural disasters and political campaigns. During the highly destructive 2010 Haiti earthquake, for instance, thousands of ordinary citizens around the world self-organized to aggregate, translate, and plot social media messages of distress to support the disaster response [2]. Similarly, the events of the 2011 Arab Spring were greatly facilitated by demonstrators' access to social media [3].

It is therefore not surprising that social media has galvanized a number of research streams in domains as diverse as information systems, crisis informatics, computer science, political science, and epidemiology. One area of research focuses on the role of online behavior in offline outcomes such as depression and influenza spread [4], and political participation and opinion polarization [5], among others. Another stream investigates the factors responsible for social media engagement, especially in the context of product and brand loyalty signaling and review behavior [6]. A third branch seeks to explain what makes social media posts go viral and how they propagate through the social network via information diffusion [7]. There is, however, a dearth of research examining the confluence of exogenous shocks such as traumatic events and the social media response to them in the production of a particular stance outside political and religious contexts.

Filling this gap in the literature is a matter of scholarly urgency, given the growing demands for a more just and equal society, since it can shed light on the mechanism of cultivating an antiracist attitude. Therefore, the research question at the core of this study is: *What are the social media factors influencing the development of a sense of justice?*

To answer this question, we integrate Paulo Freire's theory of critical consciousness with social capital and sociolinguistic theory to provide a model of racial justice sense development. Extending the phenomenological model of Landreman et al. (2007), we show how exposure to diversity combined with self-reflection and a critical incident (such as George Floyd's murder) jointly contribute to the development of critical consciousness regarding race. This study contributes to the information systems (IS) literature in three ways. First, to our knowledge, this is the first study in the IS

domain to integrate critical consciousness, social capital, and sociolinguistic theories for the purpose of critically analyzing the mechanism of racial justice development in online spaces. Second, we demonstrate the connection between a person's linguistic patterns and social network, on one hand, and their likelihood of participating in online racial justice discourse, on the other. Third, we illustrate the difference between two types of social network impact on racial justice development—the users one follows vs. one's own followers—and show that the size of the former group has a significant positive effect on one's likelihood of engaging in social justice discourse.

The remainder of this paper is organized as follows. In Section 2, we present the theoretical framework informing this study and advance testable hypotheses. Section 3 describes the data and methodology used to test the hypotheses. Section 4 presents the results of our analysis, and in Section 5 we offer a detailed discussion and implications. Section 6 concludes the study with a summary of the key findings, limitations, and possible directions for future research.

2. Literature Review and Hypotheses

The death of George Floyd sparked debates and collective action, transcending the tragedy of a family facing a personal loss and energizing a global discussion of systemic racism. While members of social groups experiencing discrimination cannot avoid facing their disparity with the dominant group, representatives of the dominant group are not forced to reckon with the reasons for, nor consequences of, the unjust social system. The choice exclusive to the dominant group of whether to engage with inequality or to ignore it embodies the difference between naïve and critical consciousness. The former is the hallmark of individuals remaining ignorant of the social justice issues characterizing their community, while the latter requires actively seeking to understand the structural problems plaguing society [8]. Developed by Freire and taken up by philosophers, sociologists, psychologists, and educators, the theory of critical consciousness presents a framework for the transition from naïve to critical engagement with social structures which can be particularly useful in unpacking the mechanism of developing a racial justice attitude online.

2.1. Critical Consciousness

Critical consciousness entails "learning to perceive social, political, and economic contradictions, and to take action against the oppressive elements of reality" [9]. In particular, it involves a dual process of reflection and action geared toward transforming social systems

and conditions [10]. Although often used interchangeably, critical consciousness differs from empowerment due to its strong commitment to liberatory practices and active pursuit of justice not just for the focal individual but for other oppressed individuals or groups as well [11]. These features enable critical consciousness to develop a type of critical literacy in the individual that lets them "read the world" and the systems at work in it [12]. Critical literacy has made the theory of critical consciousness a preferred framework for analyzing race-based oppression in the U.S. [13].

In the realm of IS research, critical consciousness has been employed to build critical data literacy which adds an emancipatory dimension to the traditional definition of data literacy as a technical ability to engage with data [14]. It has recently also been used as part of the process of developing critical data visualization skills [15], but extant research in the field has been at the conceptual level. We seek to leverage the framework of critical consciousness by extending a model proposed by Landreman et al. (2007) in the context of university education. Specifically, through in-depth interviews of university educators from 14 different racial and ethnic backgrounds, the authors develop a phenomenological model of the mechanism through which the transition from naïve to critical consciousness occurs. According to the model, attaining critical consciousness is predicated on the joint influence of a tripartite system of factors that includes diversity exposure, ability for self-reflection, and a critical incident [16].

Exposure to diversity consists of having lived experiences with people whose cultural background differs from one's own culture, whereas a critical incident is a significant event, interaction, or experience serving as a catalyst for sense-making and seeking engagement with a social issue; finally, self-reflection captures attempts to understand events or circumstances and often involves re-examination of one's place in their family, school, workplace, community, or the world in general [16]. Developing critical consciousness therefore includes exposure to diverse cultures and a critical event, which together prompt the individual to reflect on the issue at hand. To extend this model to the social media context, we use the theories of social capital and sociolinguistics.

2.2. Social Capital

Social capital is an umbrella term for the benefits that accrue to an individual through their social interactions [17]. Like other forms of capital, social capital is productive in that it makes possible certain outcomes that would not have been obtained in its absence. Unlike other types of capital, however, it

resides in the structure of relations between and among actors and is therefore not the property of any single actor [17]. Additionally, social capital is not a monumental concept, exhibiting instead multiple dimensions depending on the obligations and expectations, information channels, and social norms of the social network in question [17].

One dimension considered critical in explaining human behavior pertains to the distinction between bridging and bonding social capital [18]. The difference between these two types of social capital lies in the level of engagement between actors and the variability of their background. Specifically, bridging social capital is outward-looking, inclusive, and is observed in social situations where diverse, socially heterogeneous groups of people are connected; bonding social capital, on the other hand, is inward-looking and exclusive, occurring within homogenous groups of people [18].

The two types of social capital can roughly be mapped onto the weak and strong ties in one's social network, with strong ties characterized by frequent and intense interactions in close proximity, and weak ties described as infrequently occurring and lacking intensity or propinquity [19, 20]. Based on these definitions, bridging social capital can be acquired through weak ties, since this type of connection has been found to be essential in accessing non-redundant information due to the diverse, non-overlapping knowledge and background of the underlying social actors, while bonding social capital accompanies strong ties, which are shown to be important for routine activities that require effective coordination.

Far from being limited to the physical realm, social capital accumulation has been observed in online interactions as well. Research shows, for instance, that greater social connectedness on Facebook is associated with increased bridging and bonding social capital [21, 22]. The use of Instagram and Twitter has also been found to positively influence the two types of social capital [23, 24]. While it is not immediately obvious which type of social capital—bridging or bonding—is represented by connections on reciprocal social media platforms such as Facebook where both actors must agree to become "friends" for a tie to be formed, nonreciprocal platforms like Twitter and Instagram where one actor can "follow" another even if the latter does not follow her in return, have been studied for the specific purpose of making this distinction [25].

In particular, Hofer and Aubert (2013) demonstrate the difference in social capital type accumulated through the two distinct categories of social connections on Twitter made possible by its nonreciprocal, i.e., directed, social connectedness nature: followers and followees. In their framework, followers are linked to bonding capital, since the users who have chosen to subscribe to one's updates are likely to do so based on common interests, which signals homogeneity [26]. Followees, on the other hand, represent a potentially broader and more diverse group since the focal user has the opportunity to follow film stars, world-acclaimed authors, or news organizations beyond the realm of her own contacts, and thereby gain access to nonredundant heterogeneous information [26]. Indeed, results indicate that the more followers one has on Twitter, the higher their reported bonding capital, while more followees were associated with higher perceived bridging capital [26].

Relating the two types of social capital to the diversity–event–self-reflection model advanced by Landreman et al. (2007), it is expected that bridging social capital, characterized by interactions within loosely-connected heterogeneous social groups, will have a stronger impact on developing critical consciousness than bonding social capital, which is based on homophily rather than diversity. We therefore hypothesize that:

H1: Bridging social capital will positively impact the likelihood of exhibiting a racial justice attitude online.

To account for the less diverse makeup of social relations responsible for the production and accumulation of bonding social capital, we offer two competing hypotheses regarding its effect on racial justice discourse likelihood.

H2a: Bonding social capital will positively impact the likelihood of exhibiting a racial justice attitude online, and

H2b: Bonding social capital will negatively impact the likelihood of exhibiting a racial justice attitude online.

2.3. Social Media Displays of Self-Reflection

The final element of the diversity–event–self-reflection model is self-reflection, or one's attempt to understand events or circumstances and re-examine one's place in a social group [16]. Self-reflection has been studied extensively in social psychology, where it has been established as a metacognitive trait central to self-regulation, or the control of one's behavior, thoughts, and emotions in response to external stimuli [27]. Although self-reflection is commonly measured with the self-reflection and insight self-report scale [28], recent advances in computational- and socio-linguistics have enabled the use of subjects' secondary data in the form of user-generated content on social media platforms, blogs, and other virtual communities to estimate one's capacity to reflect on oneself and the world [29].

One type of user-generated content that has been found particularly helpful in assessing a subject's thinking style, including self-reflection, attentional focus, and emotionality, among others, is user-generated text [29]. In particular, the use of certain parts of speech (specifically, nouns and prepositions), or types of words indicating causal and insight language (e.g., because, hence; think, know, consider) can indicate an active process of self-reflection [29]. The explanatory power of written text has been detected even on microblogging platforms with a character limit producing posts no longer than a single sentence [30]. Despite the shortness of individual microblogging messages, research shows that analyzing user posts over time can provide valuable insights into the user's intellectual tendencies by capturing the trends in topics covered by the messages [31]. How many topics a user is interested in and posts about serves as a proxy of intellectual breadth and may signal open-mindedness and willingness to learn—facets important to self-reflection [32]. We therefore hypothesize that:

H3: Topical range will positively impact the likelihood of exhibiting a racial justice attitude online.

3. Methodology

3.1. Data

To test our hypotheses, we collected data from Twitter. The choice of Twitter as the platform of analysis is based on two reasons: (1) Twitter's social network structure enables us to differentiate between followers and followees, and (2) Twitter's API provides ready access to rich social media data, including a user's profile description and full list of publicly available tweets over a period of time. Selecting an appropriate sample of Twitter users whose attitudes toward racial justice we would like to examine can proceed along two different paths: (1) specifying explicit search term criteria such as hashtags including "JusticeforGeorgeFloyd", "ICantBreathe, "BlackLivesMatter", and "BLM", among others, to construct the positive outcome sample, and then identifying a way to randomly select a negative outcome sample, or (2) selecting Twitter users at random, based on an unrelated criterion. Due to the potential for selection bias when applying explicit search term rules, we chose the second way of gathering a sample of users to be analyzed. Specifically, we focused on Twitter users whose tweets were captured by Twitter's COVID-19 research stream within a month of George Floyd's death (May 25 to June 22). We retained U.S.-based geo-tagged tweets only, which resulted in an initial sample of 21,471 unique users. Next, we manually cleaned the list of users, filtering out business accounts, organizations, and aliases, since the aim of the study is to uncover the factors accounting for individual critical consciousness development. This process reduced the sample to 10,778 users. We then collected each user's full set of tweets posted between May 25 and June 22. This procedure resulted in a total of 946,533 tweets. To identify tweets exhibiting a social justice attitude without imposing explicit criteria, we trained a Latent Dirichlet Allocation (LDA) topic modeling model using Python's Gensim package. LDA is a fully generative unsupervised machine learning algorithm for discovering implicit themes (topics) in large and often unstructured collections of documents. Specifically, it represents documents (tweets) as random mixtures over latent topics, where each topic is characterized by a distribution over words [33]. The algorithm requires three inputs to initiate the modeling process: a number of latent topics, a pre-processed collection of tweets (corpus), and a dictionary representation of all unique words (vocabulary), where the occurrence of each word in the corpus is represented by a count. Although the number of existing topics cannot be known with certainty a priori, it can be approximated by measuring LDA model performance over various numbers of topics. Figure 1 illustrates model performance for a number of topics ranging from 1-100.
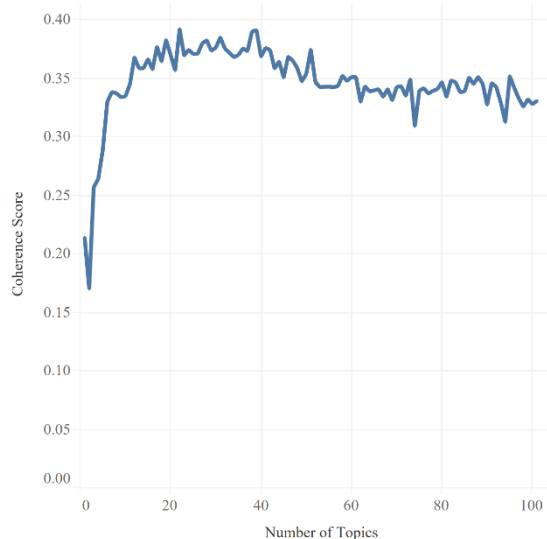

Figure 1. Coherence score for number of topics 1-100

The optimal number of topics is the one associated with the highest coherence score, where coherence evaluates the interpretability of generated topics by estimating pairwise word comparisons. Although there are different measures of coherence, we used Cv coherence, which calculates how often top words in a topic co-occur in the corpus, due to its high correlation

with human judgment of phrase interpretability [34]. The highest coherence score in Figure 1, 0.392, corresponds to an LDA model with 22 topics. We therefore used this number of topics as a model input and trained the model on the full set of 946,533 tweets over 1,000 iterations. The coherence score of the trained model is 0.4601, indicating that the additional passes over the corpus significantly improved the resulting topics' interpretability. Next, two information systems scholars and a professor of philosophy and ethics independently reviewed the top 100 words comprising each of the 22 topics produced by the trained model to identify topics related to racial justice. All three experts selected two topics in particular. Figures 2 and 3 show the word clouds associated with these topics, hereafter referred to as "black-life" and "police-protest," respectively, using the top two most frequently occurring words per topic as labels.
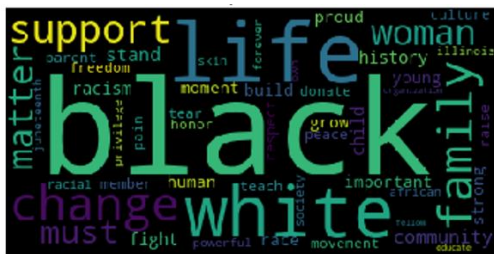

Figure 2. Top 50 words for Topic #18

Figure 3. Top 50 words for Topic #20

In the next step, we passed the corpus of 946,533 tweets through the trained model to identify the dominant topic of each tweet (since a tweet can be a mixture of several topics), where the dominant topic is the one with the highest probability weight. In the case of two or more topics of equal weight, the tweet label would be marked as "n/a" and manually resolved by the three experts. This procedure yielded a total of 12,156 tweets labeled as predominantly black-life or police-protest, corresponding to 3,229 unique Twitter users.

Since the study focuses on the transition from naïve to critical consciousness prompted by the death of George Floyd, social network diversity, and self-reflection, we are interested in Twitter users not engaged in racial justice discourse prior to the critical incident of

Floyd's murder. We therefore collected the full set of tweets posted by all 10,778 users in the month before the critical event, i.e., April 26 to May 24, 2020, and labeled them depending on their dominant topic following the procedure described above. This step identified 2,841 tweets about black-life or police-protest and led to the exclusion of 1,382 users, 478 of whom were part of the 3,229 users posting on black-life or police-protest after George Floyd's death.

Thus, the sample of users considered for further analysis consists of 9,396 users (10,778-1,382), of whom 2,751 (3,229-478) engaged in racial justice discourse following the killing of George Floyd. lthough the natural split between racial justice discourse participants (N=2,751) and non-participants (N=6,645) is a viable sample approach, we introduced two additional criteria that helped us construct a matching control sample of non-participants with demographic and social media engagement characteristics similar to those of discourse participants. Specifically, we used coarsened exact matching (CEM) to find non-participant matches from the same county as a focal participant (based on the Federal Information Processing Standard code, or FIPS, derived from the user tweets' geotag) and with approximately the same number of tweets posted to date. This procedure enabled us to control for geographic location and social media engagement, thereby fleshing out the effect of social network diversity and self-reflection on racial justice discourse likelihood. Despite the proliferation of statistical matching techniques, we adopted CEM due to its superior performance compared to older approaches [35]. In particular, CEM is a member of the monotonic imbalance bounding matching techniques and divides matching criteria into "bins," so a control entity is only marked as a match if it falls in the same bin as the focal entity along all specified criteria, i.e., covariates, where (1) the performance of each covariate measured as the reduction in distance between a treatment and a control case is independent from the performance of the other covariates, and (2) exact matching, as opposed to approximate matching, is possible—two significant improvements upon equal percent bias reducing methods such as propensity score matching [35]. The CEM procedure was carried out by means of the –cem– command in Stata MP14, where, to be admitted as a control, a racial justice discourse non-participant's tweets would have to originate in the same county as those of a discourse participant, and her total number of tweets since joining the platform would have to fall into the same octile bin as that of a discourse participant. Starting with 2,751 discourse participants and 6,645 potential control non-participants, this process found a total of 3,642 non-participant matches for 1,723

participants. The final matched sample thus includes 5,365 Twitter users.

3.2. Operationalization of the Variables

3.2.1. Dependent Variable. The three hypotheses proposed in Sections 2.2 and 2.3 aim to test whether the diversity of an online user's social media social network and her exhibited self-reflection capacity significantly impact the user's likelihood of engaging in online racial justice discourse. This likelihood is captured with a dichotomous variable that takes on the value of 1 if the user posted a tweet on the topics of black-life or police-protest in the month following George Floyd's death, and 0 otherwise.

3.2.2. Independent Variables. To extend the diversity–event–self-reflection model to a social media platform context, we adopt the following predictor variables. First, in the case of a critical event, we consider the death of George Floyd as a watershed moment in racial justice discourse and, therefore, as a viable critical event.

Second, in terms of diversity exposure, as noted in Section 2.2, social capital, with its two types, bridging and bonding, is an apt paradigm for measuring the diversity of a social network [18]. In particular, the two types of users composing a focal user's social network on Twitter, i.e., those she follows (followees) and those she is followed by (followers), have been shown to reliably predict bridging and bonding social capital, respectively [26]. Therefore, to test our hypotheses, we operationalize exposure to diversity through two measures: the focal user's number of Twitter followees to capture bridging social capital, and the focal user's number of Twitter followers to reflect her bonding social capital.

Finally, to assess self-reflection capacity in online settings, we rely on the well-established link between written language and intellectual and emotional breadth [36]. Following the review of the sociolinguistic literature in Section 2.3, we capture self-reflection capacity through the variety of topics one's tweets exhibit, measured as the number of dominant topics reflected in one's tweets posted before George Floyd's death. We lag this indicator in order to avoid confounding the critical event's influence on critical consciousness development with one's pre-incident manifestation of self-reflection. To establish the number of topics a user has engaged in, we use the labels derived from the trained LDA model in Section 3.1. Thus, topic range can theoretically take on values between 1, if user posts center on a single topic, and 22, if the user has posted tweets engaging each of the latent topics identified by the model.

3.2.3. Control Variables. Despite controlling the variation in racial justice discourse engagement likelihood attributable to geographic location and number of tweets through CEM, we introduce several user-level controls known to have an effect on Twitter-related outcomes.

Hashtag and mention usage, for instance, have been found to have explanatory power in identifying Twitter users' political affiliation and interests [37], which can have an impact on the likelihood of engaging in racial justice discourse. Additionally, embedding multimedia, such as a photo, in a tweet has been shown to increase the audience of a tweet due to the easy consumption of the tweet's content [38]. Therefore, photo-embedding behavior can be assumed to have an opportunistic dimension that could also play a role in the likelihood of engaging in racial justice discussions. We operationalize these three variables by counting the number of pre-event tweets per user containing hashtags, mentions, and photos, respectively, and dividing these instances by the total number of pre-event tweets posted by each user.

Whether a user has chosen to include a profile photo and a background photo in their profile and provide a description or a biographical note, on the other hand, have all been linked to the construction of positive self-representation, which can also have an impact on the type of discourse a Twitter user chooses to engage in [39]. To capture these three user-profile factors, we construct three indicator variables that take on the value of 1 if the user has provided a description, uploaded a background image, or posted a profile image different from the Twitter-assigned stock image, respectively, and 0 otherwise.

3.3. Model Specification

To test our hypotheses, we use ordinary least square (OLS) regression with a binary outcome, also known as a linear probability model (LPM). This model specification produces readily interpretable estimates and can be represented mathematically as follows:

$$y_{it} = \beta_0 + BX_{it} + \varepsilon_{it,}$$

where y_{it} signifies whether user t in stratum i has posted on racial justice or not, and X_{it} is a matrix of predictors. To control for the effects of geographical location and number of tweets, we construct an indicator variable based on the coarsened exact matching (CEM) strata ID and include it in the model.

As a robustness check, we follow the example of Roh et al. (2016) and use the matching strata ID as a condition group to fit a conditional logit model—another common method for evaluating binary outcomes with matched samples [40].

The conditional logit model allows us to estimate the likelihood of a Twitter user engaging in racial justice discourse compared to other Twitter users in the same county and with roughly the same total number of tweets. Since the effect of these two characteristics (geographic location and total number of tweets) on the outcome is canceled out by the conditional grouping, we are able to focus specifically on the impact of the three predictor variables on racial justice discourse participation likelihood. The model specification is as follows. Let $i = 1, 2, \ldots, n$ denote the number of strata in the sample and let $t = 1, 2, \ldots, T_i$ denote users in the ith stratum. Let y_{it} be the dependent variable taking on a value of 1 if user t in stratum i has posted a racial justice tweet and 0 otherwise. Let $y_i = (y_{i1}, \ldots, y_{iT_i})$ be the racial justice tweet outcomes for the ith stratum. Let x_{it} be a row vector of covariates. Let

$$k_{1i} = \sum_{T=1}^{T_i} y_{it}$$

be the observed number of racial justice tweets in the ith stratum. There are thus k_{1i} Twitter users participating in racial justice discourse matched to $k_{2i} = T_i - k_{1i}$ control users in the ith stratum.

Prior to the econometric analysis, we performed several checks for outliers. First, to alleviate the bias potentially caused by the presence of extreme values in the sample, all continuous variables were winsorized at 0.5% and 99.5%. To check for collinearity, we examined the variance inflation factor (VIF) of each predictor. The values of VIF ranged between 1.01 and 1.21, and had a mean value of 1.07, well within the recommended limit of 10. Table 1 shows the descriptive statistics for the variables used in the model. Due to the skewed distributions of the three predictor variables, we log-transform them before including them in the model.

Table 1. Descriptive statistics and pairwise correlations; N=5365

		Mean	S. D.	Min	Max	1	2	3	4	5	6	7	8	9
1	RACIAL_JUSTICE_DISC	0.29	0.45	0	1									
2	DEFAULT_AVATAR	0.00	0.04	0	1	-0.02								
3	BACKGRPUND_IMAGE	0.94	0.23	0	1	0	-0.03							
4	USER_DESCRIPTION	0.90	0.30	0	1	0.10*	-0.09*	0.02						
5	%PHOTO_TWEETS	0.06	0.13	0	1	0.03*	-0.02	-0.03	0.05*					
6	%HASHTAG_TWEETS	0.36	0.33	0	1	-0.03	-0.01	0.02	0	0.01				
7	%MENTION_TWEETS	0.32	0.31	0	1	0.11*	-0.02	0.02	0.08*	0.08*	-0.13*			
8	#FOLLOWEES	955.69	1462.37	6	12404	0.11*	-0.01	0.10*	0.13*	0.05*	-0.01	0.08*		
9	#FOLLOWERS	1884.55	6942.90	5	69208	0.05*	-0.01	0.04*	0.07*	0.02	0.01	0.05*	0.37*	
10	TOPIC_RANGE	3.57	1.94	1	16	0.27*	-0.02	0.04*	0.08*	0.04*	-0.09*	0.08*	0.19*	0.10*

4. Results

The proposed hypotheses were tested by fitting an LPM and a conditional logit model in Stata 13MP, with robust standard errors. Table 2 summarizes the results.

Our analysis shows that the hypotheses advanced in Section 2 are largely supported and, therefore, attest to the salience of using the diversity–event–self-reflection model of critical consciousness development in the context of social media interactions. In particular, as Table 2 shows, the number of Twitter users a focal user is following—her followees—has a significant positive impact on the likelihood of the user engaging in racial justice discourse by posting tweets related to the black-life or police-protest topics. Hypothesis 1 is thus supported, suggesting that bridging capital is indeed related to online racial justice discourse.

The next set of hypotheses is formulated as a pair of competing hypotheses because although the social capital literature has found bonding social capital to be an important determinant of positive social network outcomes, in the case of critical consciousness it represents homogeneity instead of heterogeneity. Since heterogeneity is the basis for diversity exposure in our model, bonding social capital's impact on the likelihood of engaging in racial justice discourse online is ambiguous. However, our analysis shows that although its effect across both models is smaller than that of bridging social capital, bonding social capital also contributes positively to participating in racial justice discourse online. H2a is thus supported over and against H2b, which predicted a negative impact.

Lastly, H3 states that the range of topics reflected in one's tweets will positively influence her likelihood of posting on the topic of racial justice. As the coefficients of topic range in Table 2 show, this hypothesis is strongly supported across both models, and this predictor has a stronger effect on racial justice discourse engagement than both the number of followees and followers.

Table 2. Models of racial justice discourse likelihood

Variables	(LPM)	(Cond. Logit)
DEFAULT_AVATAR	0.034	-0.221
	(0.157)	(1.209)
BACKGRPUND_IMAGE	-0.102***	-0.581***
	(0.03)	(0.168)
USER_DESCRIPTION	0.082***	0.545***
	(0.021)	(0.127)
%PHOTO_TWEETS	0.014	0.053
	(0.044)	(0.234)
%HASHTAG_TWEETS	-0.059***	-0.302***
	(0.021)	(0.109)
%MENTION_TWEETS	0.047**	0.205*
	(0.023)	(0.124)
#FOLLOWEES	0.023***	0.121***
	(0.008)	(0.035)
#FOLLOWERS	0.016**	0.056**
	(0.007)	(0.027)
TOPIC_RANGE	0.184***	0.946***
	(0.013)	(0.07)
Constant	0.199	
	(0.374)	
CEM Strata ID dummies	yes	
N	5,365	5,365
R-squared	0.18	
Pseudo R-squared		0.1

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

5. Discussion

Results of this study indicate that bridging social capital, bonding social capital, and topic range as a proxy for self-reflection capacity positively impact the likelihood of engaging in racial justice discourse online. Topic range is associated with the highest positive effect not just among the three indicators but also when compared to the effect sizes of other Twitter user features known to influence online behavior such as the use of photos, hashtags, or mentions in tweets and the level of user profile customization captured by whether a user has provided a description on their profile and chosen to upload a profile picture or a background image. This finding is important because it suggests that intellectual curiosity, implied in the breadth of topical content, is an important antecedent for critical consciousness development.

The number of accounts a Twitter user follows—her followees—have the second largest effect size among the three predictors, albeit significantly smaller than that of topic range. Twitter's unique social network mechanism, which makes it easy for users to follow others regardless of status, has been shown to result in higher bridging capital compared to other social media platforms like Facebook, Instagram, and Snapchat [23]. The ease of diverse weak tie creation allows for the discovery and diffusion of new information [41], which can quickly point a user's attention to a social issue of growing importance, as seen in the case of George Floyd's death. The big difference in effect size vis-à-vis topic range could be explained with the idea that it is more important who one follows rather than how many users she follows. The quality over quantity adage may apply in the context of critical consciousness development because although Twitter provides the ability to follow people from different walks of life and social status, there is still a strong agency component to deciding who to follow, and this decision may often be based on homophily [42].

Contrary to expectation, bonding social capital reflected in the number of followers does show a positive effect on the likelihood of engaging in racial justice discourse online. Although more closely related to strong ties and homogeneous groups, i.e., the opposite of diversity, it appears to play a role in critical consciousness development. One possible explanation has to do with the need online users have to maintain a positive self-image and be perceived by others as just and fair [43]. Since followers are considered one's direct audience and support network [26], it can be expected that the larger this community, the more social pressure the focal user would experience to appear virtuous.

An unexpected finding in this study is the negative effect the presence of hashtags and background images have on the likelihood of engaging in online racial justice discourse. The proportion of hashtagged tweets was calculated based on tweets posted in the month prior to George Floyd's death to at least partially account for endogeneity. This suggests that users with a higher proportion of hashtagged tweets in the month of May were less likely to post on racial justice issues in June. A post-hoc analysis reveals that 81 percent of the Twitter users with the highest proportion of hashtagged tweets (between 80 and 100 percent of their April tweets) posted on two dominant topics in particular. Both topics have to do with sharing personal experiences and asking for advice, such as, "what is your opinion of people who leave their mask down when no one is nearby? i walk early in the morning when barely anyone is out. and i raise it back up when i see someone approaching." The top five hashtags in the dataset are: #realdonaldtrump, #youtube, #poshmarkapp, #googlenews, and #thehill.

Together, these findings point to the salience of analyzing social media user engagement in determining individuals' likelihood of developing a critical consciousness with respect to social justice. Even though critical consciousness models are widely used to explain the formation of an antiracist attitude in adult education and social psychology, their application in the IS literature is still limited.

6. Conclusion

In this research, we examine the effect of bridging and bonding social capital and self-reflection capacity signaled through one's topical range of tweets on the likelihood of engaging in online racial justice discourse. Our results suggest that topic range is the strongest predictor of this likelihood, followed by the number of followees and followers, which capture bridging and bonding social capital, respectively. Our findings also suggest that higher proportions of hashtagged messaging, as well as uploading a background image, are negatively associated with racial justice attitude development.

This study is not without limitations. First, even though the number of followees and followers is closely related to bridging and bonding social capital, the issue of tie quality rather than quantity can only be resolved through analyzing the tweet content of focal users' first degree connections. Self-reflection can also be operationalized differently, for instance by implementing Pennebaker et al.'s (2001) linguistic inquiry and word count (LIWC) technique to provide a more granular measure of thinking style and self-reflection [44]. Another limitation of the study has to do with the cross-sectional nature of the data.

Future research can consider modeling the dynamic properties of both tweet content and one's social connections on the platform. Gender, race, and generational differences among social media users should also be interrogated as they may act as important boundary conditions in the context of online racial justice discourse. Additionally, this study treats racial justice development as a monolith, whereas Landreman et al.'s (2007) model accounts for two distinct phases of development. Future work can adapt this more granular approach to the social media context and investigate whether online racial justice attitudes exhibit stages such as those present in moral development.

Transitioning from naïve to critical consciousness is an often unsettling and painful journey, a journey feminist philosopher María Lugones calls a "pilgrimage" through various levels of comprehension and incomprehension, of intimacy and solidarity [45]. Faced with persistent calls for a more just and equal society, IS researchers have a unique opportunity to uncover and amplify the factors aiding this pilgrimage and expose those that impede it.

7. Acknowledgments

8. References

[1] S. E. Cohen, "Sandy marked a shift for social media use in disasters," *Emergency Management,* 2013.

[2] J. L. Chan, G. Nam, A. G. Marshall, and H. Purohit, "Trends in Humanitarian Health Information during 2010 Haiti Earthquake: Motivation for Curating Domain Knowledge Base," in *ISCRAM*, 2019.

[3] A. Bruns, T. Highfield, and J. Burgess, "The Arab Spring and social media audiences: English and Arabic Twitter users and their networks," *American behavioral scientist,* vol. 57, no. 7, pp. 871-898, 2013.

[4] M. Santillana, A. T. Nguyen, M. Dredze, M. J. Paul, E. O. Nsoesie, and J. S. Brownstein, "Combining search, social media, and traditional data sources to improve influenza surveillance," *PLoS Comput Biol,* vol. 11, no. 10, p. e1004513, 2015.

[5] J. A. Tucker *et al.*, "Social media, political polarization, and political disinformation: A review of the scientific literature," *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018),* 2018.

[6] L. Dessart, "Social media engagement: a model of antecedents and relational outcomes," *Journal of Marketing Management,* vol. 33, no. 5-6, pp. 375-399, 2017.

[7] S. Stieglitz and L. Dang-Xuan, "Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior," *Journal of management information systems,* vol. 29, no. 4, pp. 217-248, 2013.

[8] W. A. Smith, "The Meaning of Conscientizacao: The Goal of Paulo Freire's Pedagogy," 1976.

[9] P. Freire, *Education for critical consciousness.* Bloomsbury Publishing, 1973.

[10] P. Freire, "Pedagogy of the oppressed (MB Ramos, Trans.)," *New York: Continuum,* vol. 2007, 1970.

[11] B. D. Christens, L. T. Winn, and A. M. Duke, "Empowerment and critical consciousness: A conceptual cross-fertilization," *Adolescent Research Review,* vol. 1, no. 1, pp. 15-27, 2016.

[12] P. Mayo, "Critical literacy and emancipatory politics: The work of Paulo Freire," *International Journal of Educational Development,* vol. 15, no. 4, pp. 363-379, 1995.

[13] G. Moane, "Sociopolitical development and political activism: Synergies between feminist and liberation psychology," *Psychology of Women Quarterly,* vol. 34, no. 4, pp. 521-529, 2010.

[14] A. F. Tygel and R. Kirsch, "Contributions of Paulo Freire for a critical data literacy: A popular education approach," *The Journal of Community Informatics,* vol. 12, no. 3, 2016.

[15] C. D'Ignazio and R. Bhargava, "13. Data visualization literacy: A feminist starting point," *Data Visualization in society,* p. 207, 2020.

[16] L. M. Landreman, C. J. Rasmussen, P. M. King, and C. X. Jiang, "A phenomenological study of the development of university educators' critical consciousness," *Journal of College Student Development,* vol. 48, no. 3, pp. 275-296, 2007.

[17] J. S. Coleman, "Social capital in the creation of human capital," *American journal of sociology,* vol. 94, pp. S95-S120, 1988.

[18] R. D. Putnam, "Bowling alone: America's declining social capital," in *Culture and politics*: Springer, 2000, pp. 223-234.

[19] N. E. Friedkin, "Information flow through strong and weak ties in intraorganizational social networks," *Social networks,* vol. 3, no. 4, pp. 273-285, 1982.

[20] M. S. Granovetter, "The strength of weak ties," in *Social networks*: Elsevier, 1977, pp. 347-367.

[21] M. Burke, R. Kraut, and C. Marlow, "Social capital on Facebook: Differentiating uses and users," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011, pp. 571-580.

[22] N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of Facebook "friends:" Social capital and college students' use of online social network sites," *Journal of computer-mediated communication,* vol. 12, no. 4, pp. 1143-1168, 2007.

[23] J. Phua, S. V. Jin, and J. J. Kim, "Uses and gratifications of social networking sites for bridging and bonding social capital: A comparison of Facebook, Twitter, Instagram, and Snapchat," *Comput. Hum. Behav.,* vol. 72, pp. 115-122, 2017.

[24] M. Kaigo, "Social media usage during disasters and social capital: Twitter and the Great East Japan earthquake," *Keio Communication Review,* vol. 34, no. 1, pp. 19-35, 2012.

[25] A. E. Marwick and D. Boyd, "I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience," *New media & society,* vol. 13, no. 1, pp. 114-133, 2011.

[26] M. Hofer and V. Aubert, "Perceived bridging and bonding social capital on Twitter: Differentiating between followers and followees," *Computers in Human Behavior,* vol. 29, no. 6, pp. 2134-2142, 2013.

[27] P. J. Silvia and A. G. Phillips, "Evaluating self-reflection and insight as self-conscious traits," *Personality and Individual Differences,* vol. 50, no. 2, pp. 234-237, 2011.

[28] A. M. Grant, J. Franklin, and P. Langford, "The self-reflection and insight scale: A new measure of private self-consciousness," *Social Behavior and Personality: an international journal,* vol. 30, no. 8, pp. 821-835, 2002.

[29] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of language and social psychology,* vol. 29, no. 1, pp. 24-54, 2010.

[30] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in Twitter," in *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, 2014, pp. 51-60.

[31] K. Sasaki, T. Yoshikawa, and T. Furuhashi, "Twitter-TTM: An efficient online topic modeling for Twitter considering dynamics of user interests and topic trends," in *2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS)*, 2014: IEEE, pp. 440-445.

[32] M. Zappavigna, *Discourse of Twitter and social media: How we use language to create affiliation on the web*. A&C Black, 2012.

[33] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research,* vol. 3, no. Jan, pp. 993-1022, 2003.

[34] M. Röder, A. Both, and A. Hinneburg, "Exploring the space of topic coherence measures," in *Proceedings of the eighth ACM international conference on Web search and data mining*, 2015: ACM, pp. 399-408.

[35] S. M. Iacus, G. King, and G. Porro, "Causal inference without balance checking: Coarsened exact matching," *Political analysis,* vol. 20, no. 1, pp. 1-24, 2012.

[36] V. Elizabeth, "Another string to our bow: Participant writing as research method," in *Forum qualitative Sozialforschung/Forum: qualitative social research*, 2008, vol. 9, no. 1.

[37] A. Hanna, B. Sayre, L. Bode, J. Yang, and D. Shah, "Mapping the political Twitterverse: Candidates and their followers in the midterms," in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[38] C. Boididou *et al.*, "Verifying information with multimedia content on twitter," *Multimedia Tools and Applications,* vol. 77, no. 12, pp. 15545-15571, 2018.

[39] H. Park, B. H. Reber, and M.-G. Chon, "Tweeting as health communication: health organizations' use of Twitter for health promotion and public engagement," *Journal of health communication,* vol. 21, no. 2, pp. 188-198, 2016.

[40] D. McFadden, "Conditional logit analysis of qualitative choice behavior," 1973.

[41] S.-A. A. Jin and J. Phua, "Following celebrities' tweets about brands: The impact of twitter-based electronic word-of-mouth on consumers' source credibility perception, buying intention, and social identification with celebrities," *Journal of advertising,* vol. 43, no. 2, pp. 181-195, 2014.

[42] E. L. Frederick, C. H. Lim, G. Clavio, and P. Walsh, "Why we follow: An examination of parasocial interaction and fan motivations for following athlete archetypes on Twitter," *International Journal of Sport Communication,* vol. 5, no. 4, pp. 481-502, 2012.

[43] E. T. Panek, Y. Nardis, and S. Konrath, "Mirror or Megaphone?: How relationships between narcissism and social networking site use differ on Facebook and Twitter," *Comput. Hum. Behav.,* vol. 29, no. 5, pp. 2004-2012, 2013.

[44] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: LIWC 2001," *Mahway: Lawrence Erlbaum Associates,* vol. 71, no. 2001, p. 2001, 2001.

[45] M. Lugones, *Pilgrimages/peregrinajes: Theorizing coalition against multiple oppressions*. Rowman & Littlefield Publishers, 2003.