Words and Action: Modeling Linguistic Leadership in #BlackLivesMatter Communities

Dani Roytburg[1, 2], **Deborah Olorunisola**[3], **Sandeep Soni**[1], **Lauren Klein**[1, 4]
[1]Emory University, Department of Quantitative Theory & Methods
[2]Emory University, Department of Computer Science
[3]Yale University, Department of Data Science and Statistics
[4]Emory University, Department of English
{djroytburg, dolorunisola}@gmail.com, {lauren.klein, sandeep.soni}@emory.edu

Abstract

In the wake of the 2024 US presidential election, pundits on both the left and the right pointed to a conservative backlash against "woke politics" to explain the election's outcome. These politics, rooted in substantive beliefs about equity and justice–and particularly racial justice–owe their most recent rise to prominence to the Black Lives Matter (BLM) movement. A significant body of work, both qualitative and quantitative, has documented how BLM was able to move these beliefs from the margin to the mainstream. In this paper, we focus on the words that index these beliefs, devising a novel method of modeling semantic leadership across a set of communities associated with the BLM movement that is informed by domain-specific theory about Black Twitter. We describe our bespoke approaches to time-binning, community clustering, and connecting communities over time, as well as our adaptation of state-of-the-art approaches to semantic change detection and semantic leadership induction. We find evidence at scale of the leadership role of BLM activists and progressives, as well as of Black celebrities. We also find evidence of sustained conservative engagement with this discourse, suggesting an alternative explanation for how we have arrived at the present political moment.

Introduction

In early 2020, when the Black Lives Matter (BLM) movement had secured its place in the US national consciousness and when this project began, it was still necessary to provide evidence of how battles over social and political change are waged through both words and action. Writing again in April 2025, as we prepare this paper for final submission, it is evident to all (or should be) that words index larger concepts and debates. We need look no further than the US federal government, where ideas about equity and justice–and racial justice in particular–have not only become renewed topics of debate; the terms "equity" and "racial justice" themselves have been banned from government use (Yourish et al. 2025). While this act of censorship serves as a negative example, ample evidence also points to how words can open up conversations, introduce new ideas into public consciousness, and positively impact social behavior and government policy alike (Dunivin et al. 2022; Yan, Chiang, and Lin 2024). In this paper, we return to the Black Lives Matter movement, which we see as one of the most successful contemporary examples of the positive value of words (among other contributions), to explore with precision how words enter into a specific sociopolitical conversation, how they transform as that conversation evolves, and who is responsible for carrying forward those new or changed meanings.

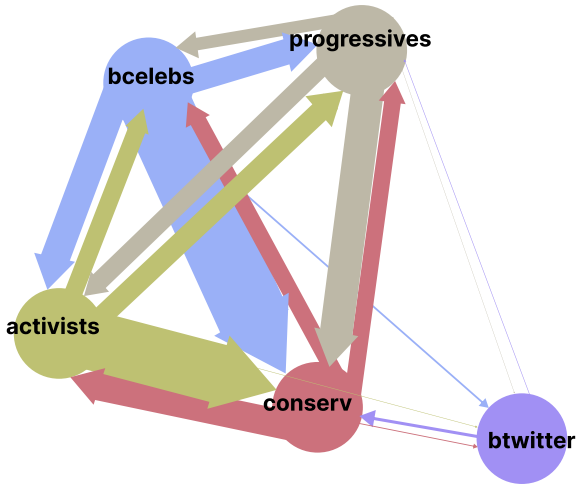

Figure 1: **Semantic leadership in #BlackLivesMatter**: Nodes represent the most central communities of the BLM network. The thickness of the edges are proportional to the number of word changes shared by each pair.

While there exists abundant research on Black Lives Matter (BLM), both as a hashtag (Choudhury et al. 2016; Jackson, Bailey, and Welles 2020; Le-Khac, Antoniak, and So 2023; Giorgi et al. 2022; Jones, Nurse, and Li 2022) and a movement (Freelon, McIlwain, and Clark 2016; del Rio 2020; Clark 2024), there are still open questions about how precisely BLM was able to "push the mainstream public sphere on issues of social progress" (Jackson, Bailey, and Welles 2020, p. xxvii). This is largely due to the unique community structure of Twitter, in which communities are defined not by individual users but by hashtags, and by the conversations among users that hashtags enable (Tufekci 2013;

Spiro and Monroy-Hernández 2016; Jackson, Bailey, and Welles 2020; Brock 2020). Operationalizing these medium-specific theories about community formation on Twitter, as well as about "Twitter time," we present a multipart model that allows us to 1) detect the latent communities associated with the #BlackLivesMatter hashtag and link them over time; and 2) detect semantic changes associated with each community using word embeddings. Together with a measure of semantic leadership (Soni, Klein, and Eisenstein 2021), we are able to induce a semantic leadership network (see Figure 1), allowing us to identify the communities responsible for introducing new or changed word meanings into the BLM network and the communities responsible for adopting (or co-opting) them.

We find that the community of BLM activists plays an outsized role in introducing semantic changes into the network, and that these changes are most consistently taken up by the conservative community. We find a similar albeit weaker signal between the progressive and conservative communities, confirming the role of left-leaning discourse in shaping the general terms of debate (Le-Khac, Antoniak, and So 2023). We also find that Black celebrities significantly shape the discourse, introducing word changes that the center/left news media as well as the conservative community later take up. The scale of our data, combined with our methods of validation, provide a new layer of evidence to support the largely qualitative scholarship that affirms the role of the BLM movement for bringing ideas about racial justice from the margin to the mainstream. Our results also provide an alternative explanation for how we have arrived at the present political moment. Contrary to a narrative of conservative backlash against a movement that became "too woke", we find that the conservative community — as the largest follower of word changes across the entire timespan of our study — had been engaging directly with the BLM movement from its very start. Today, we see the end goal of this engagement: to distort and disrupt the messaging of the BLM movement, and ultimately, to weaponize the movement's words against its most closely held beliefs.

To summarize, our contributions are as follows:

- Evidence at scale of how BLM activists shaped the discourse around racial justice, pushing new ideas from the margin to the mainstream
- Early examples of specific terms (e.g. *theory*) that have since become central to government policy and debate
- A precise network structure of the communities that comprise the #BlackLivesMatter hashtag, derived from extensive hand-labeling and domain-specific research
- An example of how domain-specific theory–here, media studies scholarship on Black Twitter–can inform technical modeling approaches

Background and Prior Work

Research on Black Twitter

The topic of Twitter, and Black Twitter in particular, has been the subject of significant research in the fields of communication and media studies, among other humanities and social science fields. In *Distributed Blackness* (2020), André Brock theorizes the community structure of Black Twitter as consisting of the people who are online and in conversation with each other at any given time. He also theorizes "Twitter time" as non-linear, defined more by specific events and the bursts of conversation they prompt than by any standard timescale. In *Hashtag Activism* (2020), Sarah J. Jackson, Moya Bailey, and Brooke Foucault Welles propose a similar conception of Twitter communities organized by the hashtag itself. Hashtags, they explain, "designate collective thoughts, ideas, arguments, and experiences that might otherwise stand alone." Additional work on Black Twitter has considered the perspectives of the users themselves (Clark 2024), its historical antecedents (Klassen et al. 2022), and its possible futures (Walcott 2024). This qualitative work constitutes the foundation for our project. Our contribution represents an attempt to formalize the concepts and structures that these media studies scholars describe.

Online Social Movements

Several characteristics of online social movements, including the motivation and role of participants (Tufekci and Wilson 2012; Bozarth and Budak 2017; Jones, Nurse, and Li 2022), activist influence on potential recruits (González-Bailón et al. 2011), formation and sustenance of collective identities (Brown et al. 2017), and mobilization of activists (Theocharis et al. 2015; Freelon, McIlwain, and Clark 2018; Mundt, Ross, and Burnett 2018; Yan, Chiang, and Lin 2024; Mendelsohn et al. 2024), have been extensively studied. BLM, in particular, has been studied as a unique example of a highly contested social movement with a sizeable online footprint (Arif, Stewart, and Starbird 2018; Stewart et al. 2017; Gallagher et al. 2018; Giorgi et al. 2022). The online discourse of this movement has been shaped by tragic offline events (Peng, Budak, and Romero 2019) and has persisted for almost a decade (Dunivin et al. 2022). In terms of our project, the closest related work is Freelon, McIlwain, and Clark's "Beyond the Hashtags" report (2016), which employs community detection followed by hand-labeling by domain experts in order to trace the changes in network structure over the first 18 months of the BLM movement. Recognizing the knowledge and expertise involved in this project, our goal was to carry forward the authors' work. Here, we extend the timeframe of analysis by another five years; develop new methods that allow us to expand the initial set of communities labeled by their team; and introduce an additional layer of linguistic analysis that complements the original study.

Linguistic methods have also been used to study different aspects of BLM. For example, the dynamics of participation have been shown to be correlated with textual features of tweets (Choudhury et al. 2016); the impact of police killings has been shown to reflect in the emotional narratives of social media participants (Field et al. 2022); hashtag use has been shown to draw attention (Blevins et al. 2019; Yan, Chiang, and Lin 2024) or push back against issues (Gallagher et al. 2018). Framing and rhetorical strategies have been shown to be means of differentiation and coalition forming (Stewart et al. 2017; Wilkins, Livingstone, and Levine

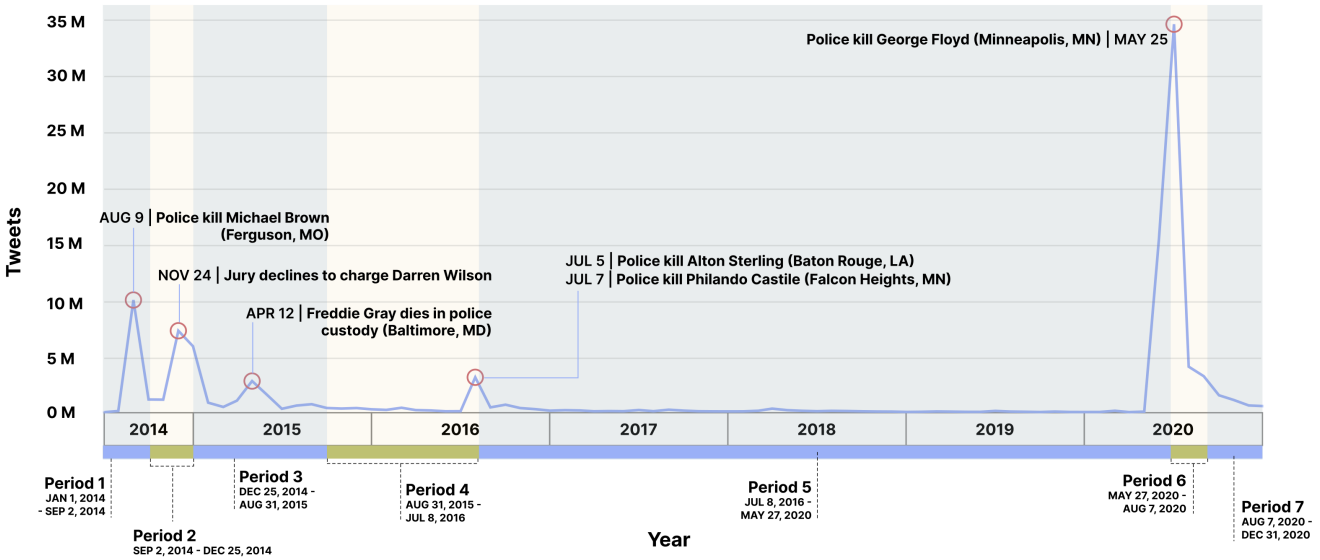

Figure 2: **Volume of tweets per month**: We visualize the temporal distribution of tweets in our dataset in the style of Freelon et al (2016). The seven temporal partitions are indicated here both by color and a reference label (e.g., Period 1). Some pivotal events to the BLM movement in each partition are annotated.

2019; Mendelsohn et al. 2024). We extend this line of work by inducing the leadership structure within the coalition of communities through the adoption of linguistic changes by community members.

Semantic Change and Leadership

Sociocultural changes often get encoded as lexical semantic changes in language (Tahmasebi, Borin, and Jatowt 2019). Consequently, computational methods to automatically extract semantic changes, given a timestamped text collection, have been developed with increasing regularity (Wijaya and Yeniterzi 2011; Kim et al. 2014; Kulkarni et al. 2015; Hamilton, Leskovec, and Jurafsky 2016b; Giulianelli, Del Tredici, and Fernández 2020; Card 2023, *interalia*). Semantic changes detected by such methods are demonstrably effective in two ways: 1) as first-order objects of analytical interest, such as in tracking the evolution of political issues (Rudolph and Blei 2018) or shift in racial attitudes over time (Garg et al. 2018); 2) as units that can be aggregated to recover the latent leadership structure among scholarly articles (Soni, Lerman, and Eisenstein 2020; Soni, Bamman, and Eisenstein 2022) or newspapers (Soni, Klein, and Eisenstein 2021). In this work, we follow the latter approach, using word embeddings to unveil the leadership dynamics between communities and their role in shaping online discourse of the BLM movement.

Data

Our dataset consists of tweets containing "#BlackLivesMatter" between June 1, 2014 and December 31, 2020. These tweets were collected in two phases: tweets between June 1, 2014 and May 31, 2015 were rehydrated in April 2021 using `twarc`[1] from tweet IDs provided by Freelon et al. (2016).[2] Our choice of this data source was motivated by our desire to build upon the hand-labeled communities in that paper, which we then used to seed our own community detection step (see next section). Tweets between June 1, 2015 and December 31, 2020, were collected using Twitter's search API between April and August 2021. Together, this resulted in a total of 82,661,364 tweets from 14,914,359 unique users.

Filtering

Our analysis proceeds in two steps: a community detection step (see Community Detection and Matching) and a semantic leadership network induction step (see Leadership Network Induction). For the former, we filtered out all tweets tagged as non-English using `langid` (Lui and Baldwin 2012); for the latter, we did additional filtering passes to remove tweets that were retweets, exact duplicates, or tagged as non-English by `langid` (Lui and Baldwin 2012) and `fasttext` (Grave et al. 2018). The first round of filtering left 69,560,572 tweets; the second round of filtering left 32,652,123 tweets. The breakdown of tweets by timebin and unique users is shown in Table 1.

Ethics and Privacy

Throughout our analysis, we remain attentive to concerns about extracting knowledge and intellectual labor from online communities for academic research (Bruckman 2002;

[1] https://twarc-project.readthedocs.io/en/latest/

[2] In addition to the #BlackLivesMatter hashtag, these tweets also included 44 additional "keywords related to BLM and police killings of Black people under questionable circumstances" (Freelon, McIlwain, and Clark 2016, p. 21)

Period	Time Span	Days	Pre-filtered Tweets	Post-filtered Tweets	Post-filtered Users
1	`Jan 01, 2014 - Sep 01, 2014`	244	7,852,863	1,809,274	331,479
2	`Sep 02, 2014 - Dec 24, 2014`	114	11,590,864	2,744,459	497,168
3	`Dec 25, 2014 - Aug 30, 2015`	249	6,775,714	1,804,867	228,935
4	`Aug 31, 2015 - Jul 07, 2016`	312	2,539,907	695,180	143,220
5	`Jul 08, 2016 - May 26, 2020`	1419	7,196,284	1,630,376	347,567
6	`May 27, 2020 - Aug 06, 2020`	72	41,593,336	5,252,117	1,659,510
7	`Aug 07, 2020 - Dec 31, 2020`	146	5,112,396	877,086	214,050

Table 1: **Descriptive statistics**: Tweet and user statistics for each temporal partition. Start and end dates are inclusive.

Jules, Summers, and Mitchell 2018; Dym and Fiesler 2020; Jackson, Bailey, and Welles 2020; Walsh 2023). These concerns motivate our modeling approach, which is designed to extend rather than replace existing scholarship. In addition, our multiracial project team has members who have engaged with the BLM movement in various roles, on and offline. With respect to concerns about privacy and attribution, we are not required to identify individual users in most cases, as the majority of our analysis is undertaken at an aggregate level. In the four instances where we identify individual users or their tweets, we take a context-specific approach (Dym and Fiesler 2020). Figure 6 requires that we give a sense of the users that comprise each community. In this case, to avoid unwanted exposure, we rank users by in-degree and name only those within the top five of each community that a) have maintained public accounts through Twitter's transition to X; and b) as of September 2024, still meet the "reasonably public" threshold of 3000 or more followers as formulated by Freelon et al. (2016). Table 2 requires that we provide examples of the word changes that our model detects. Here, we follow current best practice by paraphrasing tweets rather than quoting directly (Dym and Fiesler 2020; Walsh 2024) because we do not want to expose activist users or users with low follower counts to unwanted attention or possible harm (Jules, Summers, and Mitchell 2018). We also confirm via the current X.com search interface that our "ethically fabricated" language (Markham 2012) is not traceable back to the original user. The two users mentioned by name in the paper are public figures and as such, have upwards of one million followers.

Community Detection and Matching

Our first goal is to find relatively persistent communities of Twitter users throughout the full timespan of this study. In this section we describe the procedure for partitioning the timespan, mapping nodes to clusters within temporal partitions, and matching clusters across temporal partitions to find cohesive communities.

Defining Discrete Time Periods

We partition the dataset around the local maxima of tweet frequency, allowing irregular intervals. We find that the relative maxima correspond to real-world events relating to BLM. For instance, the first local maximum is found on August 11th, 2014, two days after the murder of Michael Brown (Anderson 2016; Freelon, McIlwain, and Clark 2016). We use these real-world events as a qualitative baseline for validating possible maxima. When additional maxima occur within 30 days of another maximum, they are grouped into one period (see Table 3). From the 11 relative maxima that were extracted using this method, six became markers for time periods resulting in seven intervals (see Figure 2 and Table 1).

Time-Specific User Clustering

User graphs We use Freelon's Twitter Subgraph Manipulator (TSM)[3] package to construct graphs from three forms of user-to-user interaction on Twitter found in tweets: replies, mentions, and retweets. These graphs are used to partition users into disjoint clusters in each of the seven temporal bins.

Clustering objective To obtain user clusters, we maximize intra-group modularity (Newman and Girvan 2004; Newman 2006) by modifying the Louvain clustering algorithm (Blondel et al. 2008) — a well-known algorithm for detection of an unconstrained number of communities — using the TSM package.

Because of our goal of extending the communities identified in Freelon, McIlwain, and Clark (2016), which were hand-labeled by experts on BLM and Black Twitter, we augment the basic Louvain approach with six seed communities, consisting of 60 users, which were reconstructed from the communities documented in the original study. This improved the overall coherence of the clusters, which initial experiments had demonstrated to be poor.

Clusters We extract the top 50 clusters for each temporal bin, which can be seen as *slices* of the larger BLM network, in the sense that they capture the time-bound representation of the BLM community structure.

From Clusters to Communities

To find persistent communities from the ephemeral clusters obtained in the previous step, we cast the problem as a *matching* problem in a bipartite graph. Specifically, for a pair of successive intervals, we construct a bipartite graph linking each cluster from one interval to every cluster in the next

[3]https://github.com/dfreelon/TSM

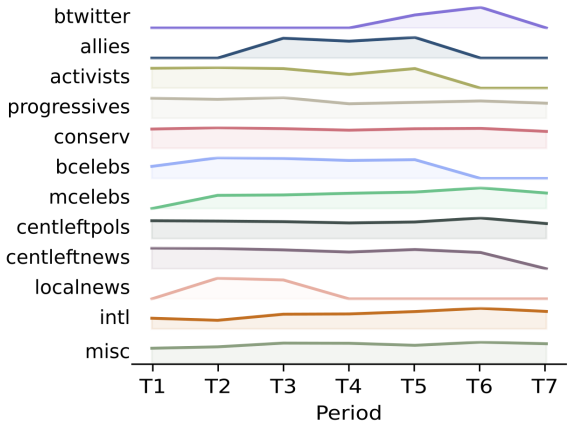

Figure 3: **Tweet counts over time** after filtering, by community. Grouping subcommunities allows for more consistent representation of each community over the entire span.

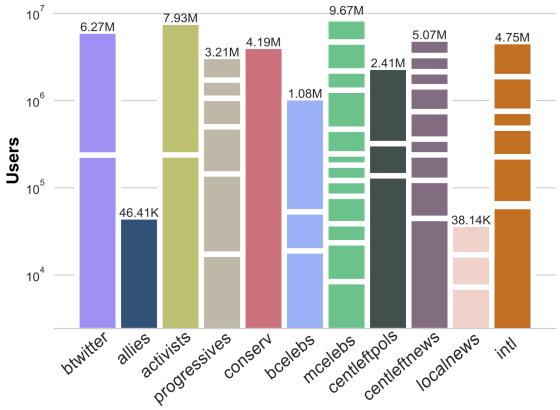

Figure 4: **Community size**: Count of unique users in each community. Bar segments represent subcommunity clusters with height proportional to cluster size.

interval. The weight on these links represents similarity between the two clusters, and is calculated as a weighted Jaccard score taking into account the in-degree of the users in a cluster (Freelon, Lynch, and Aday 2015). Following Freelon et al., we experimented with pruning the "weaker" edges if their similarity was below 0.3. However, since our intervals span over months or years, we see a greater level of fragmentation at this threshold. Consequently, we used a knee plot, which highlights the tradeoff between the number of matched communities and the jaccard threshold. Using this diagnostic, we lower the threshold to 0.07, a brightline that permits persistent communities with noise confined to spurious matches (see Figure 7).

We repeat the matching process between each successive temporal interval, and throw out any intermediary edges with similarity less than 0.07, effectively identifying the chains of clusters that remain most persistent over time. In summary, the sets of disjoint, independent clusters are matched to the preceding and succeeding periods producing 65 cohesive and persistent communities (see Figure 3).

Labeling

Labeling Users

With the sequenced slices matched, we can begin to develop qualitative characteristics for the 65 communities that result. We sought to generate a basic profile for each by surveying the top 20 users ranked by in-degree in each community, and giving each label from a set of iteratively redefined labels. The decision to use the top 20 users was heuristic; on average, the top 20 users made up 46.7% of the total in-degree, demonstrating their centrality to each of the 65 communities. A sub-label is included which qualifies the sort of membership a user has. For instance, a CNN journalist like **@donlemon** would be classified as "Established Media", sub-labeled "Journalist." Each user is given a second optional label if the first is insufficiently representative; these labels are not hierarchical, as they are jointly important. From this set of classifications and from background research, we agree upon a set of categories which could maximally encompass our communities.

In order to establish label agreement, the annotators constructed a series of pilot tests. These pilot test annotations yielded 83% agreement between annotators on average. With this established, each of the four annotators labeled 360 users. We randomly selected six communities (120 users) to be annotated by two annotators for agreement. Thus, annotators labeled 1,200 total users once and 120 twice to check for agreement. We computed Fleiss's Kappa from the random pairwise assignment. Since some users received two non-ordinal labels, we present the maximum agreement score as reference. On average, interrater agreement was 0.798. After coming to consensus on how to define the mismatched labels, we continued from this user-level analysis to consolidating labels for the communities more broadly.

Describing and Grouping Communities

To develop a stronger sense of each community as a whole, the annotators used the set of labels assigned to its top 20 users to define a short description of each community. Using the two to four alternative descriptions for each community, the annotators then agreed on a final short description. In order to conduct semantic change analysis and perform permutation tests, it is necessary to condense the number of communities analyzed. Using the final short descriptions, we merged our 65 communities into 12 larger groups (see Figure 4 and Figure 6; "miscellaneous" not shown).

We excluded two groups from this analysis, one for not being written in English and the other for containing spam. Since communities on Twitter change in membership over time, it is important to check if these group labels are apt. Once grouped into the 12 macro communities, the average in-degree of the top 20 users was no less than 20%, suggesting that, relative to the size of their respective communities,

these users play a central role in its dialogue. Thus, their presence provides important information about the background of other users and potential insight into how they may approach discussing social issues.

Leadership Network Induction

We adopted the method from Soni, Klein, and Eisenstein (2021) to induce a semantic leadership network between document sources. Soni et al. 2021 considered a set of historical newspapers as document sources and individual articles as documents. In contrast, we consider our Twitter user communities as document sources and individual tweets as documents.

Besides this superficial difference in the setup, there are two substantive differences. First, Soni et al. 2021 divided their timestamped collection of newspaper articles into temporal slices of equal timespans. In contrast, our tweet collection is divided into intervals demarcated by timestamps of real-world incidents. Second, Soni et al. 2021 considered document sources to remain fixed over time. In contrast, we consider communities of Twitter users as document sources which may undergo changes as a result of users joining, leaving and migrating between communities. We thus apply the methodology to our data but note that the interpretation of influence across communities should be qualified to account for these user dynamics.

For completeness, we briefly describe the core methodology here. Consider any tweet i as a sequence of tokens (words) $w_i = (w_{i,1}, w_{i,2}, \ldots, w_{i,n_i})$ from a finite vocabulary $\mathcal{V}$, where n_i indicates the tweet length. Our corpus consists of N tweets, $\mathcal{W} = \{w_1, w_2, \ldots, w_N\}$. Each tweet has two labels: a discrete timestamp t_i, obtained by binning the document timestamps into one of T bins; and a community label s_i based on the community of the tweet's author.

Semantic Change Detection

To identify semantic changes, we learn temporal word-type embeddings from data. Other methods to detect semantic changes using token embeddings have been proposed but with mixed success (Laicher et al. 2020). Consequently, we approach the task as learning word embeddings and then enhancing them to account for information from temporal or other facets. Words are mapped to embeddings, typically by maximizing the following probability under the skipgram language model (Mikolov et al. 2013) between any pair of nearby tokens $(w_j, w_{j'})$ in a single document ,

$$P(w_{j'} \mid w_j) \propto \exp\left(\mathbf{v}_{w_{j'}} \cdot \mathbf{u}_{w_j}\right), \tag{1}$$

where $\mathbf{v}_{w_{j'}}$ is the "output" embedding of $w_{j'}$, and $\mathbf{u}_{w_j}$ is the "input" embedding of w_j; both set of embeddings are the parameters of the skipgram model that are learned from data. If documents are timestamped, then every input embedding can be split into a core embedding and a time-specific deviation as follows,

$$\mathbf{u}_{w_{i,j}}^{(t_i)} = \mathbf{b}_{w_{i,j}} + \mathbf{r}_{w_{i,j}}^{(t_i)}, \tag{2}$$

where $\mathbf{b}_{w_{i,j}}$ is the base embedding of the word $w_{i,j}$ and $\mathbf{r}_{w_{i,j}}^{(t_i)}$ is the residual for time t_i. The residual turns the core meaning to a temporally specific meaning and are hence regularized towards zero to follow the assumption that for most words the core meaning remains intact over time. Embedding decomposition of this type has been used in other applications such as to identify geographic variation (Bamman, Dyer, and Smith 2014) or perceptual differences in meaning (Gillani and Levy 2019).

Once the base and the residual components of the input embeddings for each word are obtained, semantic changes can be found by comparing near-neighbors in the embedding space (Hamilton, Leskovec, and Jurafsky 2016a). The upshot, at the end of this step, is a ranked order list of triples, consisting of the word that has changed, the timestamp for the onset of the change, and the timestamp for the conclusion of the change.

Identifying Semantic Leaders

For each semantic change, we then score the communities that lead or lag that change. To do this, the input embeddings in Equation 3 are modified to include a residual term for the community of each token. The input embedding is rewritten,

$$\mathbf{u}_{w_{i,j}}^{(t_i,s_i)} = \mathbf{b}_{w_{i,j}} + \mathbf{r}_{w_{i,j}}^{(t_i)} + \mathbf{r}_{w_{i,j}}^{(t_i,s_i)}, \tag{3}$$

where $\mathbf{r}_{w_{i,j}}^{(t_i,s_i)}$ is the source-specific temporal deviation added to the temporal and atemporal components of the input embedding.

Next, a leadership score is calculated between a pair of communities s_1 and s_2 and for a given change (w, t_1, t_2), where $t_1 < t_2$ are the timestamps of a change in the meaning of word w, as a ratio of cross-correlation between the sources to the auto-correlation in meaning as follows,

$$\textsc{Lead}(s_1 \rightarrow s_2, w, t_1, t_2) = \frac{\mathbf{u}_w^{(t_1,s_1)} \cdot \mathbf{u}_w^{(t_2,s_2)}}{\mathbf{u}_w^{(t_1,s_2)} \cdot \mathbf{u}_w^{(t_2,s_2)}} \tag{4}$$

The tuple $(s_1 \rightarrow s_2, w, t_1, t_2)$ can be referred as a leadership event denoted by e. A higher $\textsc{Lead}(e)$ score indicates more leadership; a score of 1 corresponds to a baseline case of no leadership.

Inducing Leadership Network

We can aggregate the leadership scores for each word obtained from Equation 4 to construct a dense network between the communities. However, such a network may include changes that are, at best, spurious correlations due to random noise or structural biases such as temporal precedence of certain communities over others. To account for spurious correlations, we create $K = 100$ randomized datasets by randomly swapping word tokens between communities. In principle, randomization breaks the link between individual communities and their contextual word statistics. Consequently, lead-lag relationships retained in the randomized dataset must be attributed to either structural bias or random noise. We filter out such spurious lead-lag relationships by comparing the lead score for a leadership event in the original non-randomized dataset against a set of lead scores for the same event in each of the randomized

datasets. Any event in the final set of leadership events for further analysis satisfies the following criterion,

$$\textsc{Lead}(e) > \Phi_{.95}\left(\{\textsc{Lead}^{(k)}(e)\}_{k=1}^{K}\right), \quad (5)$$

where the function $\Phi_{.95}(S)$ selects the 95th percentile value of the set S. The final output of the entire procedure is a list of events that can be aggregated to produce a weighted directed network between the communities, with the weight on any edge indicating the number of linguistic changes for which the pair of communities are in a lead-lag relationship of statistical significance.

Results

Our approach offers insight into the changing meaning of conceptual keywords (Williams 1985) related to political activism and the BLM movement, as well as about the specific communities that helped to spread those changed meanings. Here we discuss several sets of semantic changes identified by our model and the significance of the overall network we construct. Taken together, these results affirm the validity of our model and its underlying community clustering approach, as well as the crucial work done by BLM activists to shape conversations about police violence, antiblack racism, and possible solutions to both. They also affirm the importance of Black celebrities in communicating the ideas of BLM to the general public, as well as the presence of conservative critics who push back against activist discourse — and, at times, co-opt its language. Below we discuss the findings from each phase of our methodological pipeline, building towards a conclusion about the decade-long lead-up to the current political climate in which the words of social justice and antiracism themselves have become the subject of legislative attention.

Semantic Changes

The first part of our model identifies the words that demonstrated statistically-significant semantic changes (Table 2). Beginning with general terms related to activism, we find that the term *activists* undergoes a shift in meaning between 2014-2015, following the murders of Freddie Gray and Sandra Bland, when it has no particular connection to the work of BLM — its near neighbors are terms like "detained," "targeted," and "scuffle" — to the summer of 2020, by which point the term is very clearly associated with BLM activists in particular, as indicated by near neighbors including "blackled," "movements," "bipoc," and "advocates." The term *action*, similarly, enters the dataset with no specific connotations of activism (its near neighbors are related to TV news channels) but, by the summer of 2020, has acquired specific valences of political change, as evidenced by the near neighbors of "meaningful," "tangible," "steps," and "stand." In a similar manner, the word *active* also acquires connotations specific to allyship, as indicated by near neighbors of "learning," "listening," and "unlearning," among others. (The terms *allyship* and *learning* undergo similar changes, acquiring each other as near neighbors along similar timeframes).

Terms like "listening" and "learning" point to more specific ways that BLM activists shaped the broader conversation about how to engage in activism around racial justice as well as how to contribute as an ally. But additional terms index specific issues and debates. For example, the term *abolish* begins with associations with the death penalty and, in 2020, acquires an expanded meaning that encompasses the abolition of the police — a key component of the BLM platform. The term *abolishing*, similarly, acquires the near neighbors of "defunding," "reallocating," "reforming," and "demilitarization" as the summer of 2020 approaches. We also find *theory*, as in "critical race theory," coalescing as a keyword during this time, acquiring associations with both "crt" and "ideology." Taken together, these word changes point to a clear coalescing around the discourse of racial justice that we have come to associate with the movement.

Leadership Network

Unexpected insights begin to emerge when looking at the leadership network overall. To be sure, we observe the influence of the BLM activists (Figure 5a), who claim an outsized number of word changes in comparison to most other communities. The progressive community, which sits slightly closer to the center than the activist community, can count a significant number of word changes as well (Figure 5d), and it would make sense that the pattern of the progressives would track that of the BLM activists but less strongly. What is perhaps unexpected is that the conservative community — moreso than the center/left news media — is the most substantial follower of both the activist and progressive communities. While we might expect a more gradual path from the left to the center, and then to the right, our method shows how the conservative community engaged directly with the activist and progressive communities from the start.

Another unexpected but not surprising result was the strong role of the community of Black celebrities in communicating the ideas of the BLM movement to the general public (Figure 5b). Indeed, Black celebrities claim 24.73% of all of word changes, and are followed by activists, then center/left news media, and conservatives in high proportion. This finding suggests that the direct association with the BLM movement may mark its activists for pushback — and, at times, outright aggression — by the conservative community, the celebrity status of those in the Black celebrity community may protect them from some of the conservative vitriol, and as a result, allow them to communicate the ideas of the activists to a broader public. We see this in their leadership of terms like *activism*, which transforms from the specific context of social movements to a word that itself indexes a role of social importance, as indicated by the near neighbors of "importance," "career," "societal," and "impact." We also see this in the terms that one would seem to associate with the BLM movement itself, such as *listen*. The connotations of learning and unlearning, as well as amplifying the voices of those with direct experience-as indicated by the near neighbors of "learn," "unlearn," "amplify," and "explain," are communicated to the center/left news media commmunity from the Black celebrities. In addition, we find this community influencing the community

Word	Span	Earlier Tweet	Later Tweet
theory	T3→T7	*#MikeBrown smearing follows classic a conspiracy **theory** pattern...*	*Critical Race **Theory** is an actual theory not just three words...*
active	T2→T6	*Downtown is **active** right now!*	*I'm calling for everyone, regardless of race, to get up and get **active**, and NOT BE SILENT!*
abolish	T4→T6	*Miscarriages of justice occur, especially to African Americans. This is precisely why the US should **abolish** executions!*	***Abolish** the police. Their actions have shown for DECADES that they do not value Black lives.*
allyship	T2→T6	*#AllLivesMatter is not **allyship**. It's not all lives that are in danger...*	*A great convo on **allyship**, the importance of unlearning, and the dangers of white apathy...*
learning	T1→T6	*Just **learning** the details about #MikeBrown. Heartsick for him, his family, his community, and all of society*	*Listening. **Learning**. Acting. Here is an update on the steps we are taking*

Table 2: **Semantic changes**: Examples of semantic changes with the tweets in which they appear, paraphrased from the original.

of progressives, passing along the changed meaning of *theory*, discussed above, as well as the more basic concept of *equity*. Interestingly, the conservative community at times learns directly from the Black celebrity community. In particular, the term *nonracist* is introduced by this community, and is adopted directly from the conservative community from there.

A final observation has to do with the role of the conservative community in introducing new word meanings-and not just adopting them (Figure 5c). While few of these words are conceptual keywords, their overall number is substantial. This finding confirms the consistent and persistent engagement of these ideas by the conservative community — not only in the wake of the summer of 2020, but from the movement's start. That these words are not conceptual keywords at once confirms prior findings about how conservatives — as well as white people as a group — demonstrate a lead-lag relationship with BLM activists (Le-Khac, Antoniak, and So 2023) and shows how they engaged at the level of discourse: through the wide array of words that make up a language.

Implications and Limitations

The wide range of statistically significant words in our dataset that are associated with political activism and with the BLM movement in particular, as well as the communities that they were introduced and/or adopted by, confirm the effectiveness of our approach for detecting semantic change and measuring leadership within a large and heterogeneous Twitter community. It also provides an additional layer of evidence for the valuable work of BLM activists, as well as Black celebrities, in advancing the cause of racial justice over the past decade. More recently, as Twitter has become X, and has prompted the scattering of the diverse communities who previously conversed together, we may need to consider additional adaptations to our method of detecting semantic change, just as we will need to reimagine the ways in which activists can spread their messages to a broader public and enlist others in their cause.

As far as the prominence of the conservative community in our study, we see significant implications for our understanding of the emergence of the "anti-woke" and "anti-DEI" agenda that, at the time of this writing, has entered legislation in nine US states (Ray and Gibbons 2021) and appears poised to wholly reshape government agencies and funding at the federal level. Whereas the common narrative is that this movement emerged in the aftermath of the summer of 2020, and has since been fueled by the Republican party, our findings show that there has always been direct conservative engagement with the discourse of Black Lives Matter. Moving forward, this finding may reshape the narrative of "conservative backlash" and, instead, point to a story of continual pushback and even outright aggression that met the movement from its start.

Finally, we would like to highlight a key absence in this project, which is the missing data that might otherwise document the crucial role of the young Black activists who fueled the movement from its start. These regular young people were essential in the movement's early phases, as they were in organizing the on-the-ground protests that took place in every city where a police killing took place. Qualitative researchers have documented their crucial role (Jackson, Bailey, and Welles 2020; Clark 2024), as have news reports (Xue 2020), yet their status as regular young people has meant that they lack the protections granted to celebrities or even to adults. As activists, they have been subjected to harassment and intimidation, and in some cases they have even been killed by the police themselves (Press 2019; del Rio 2020). As early as mid-2015, when Freelon et al released their initial findings on the network structure of #BLM, they found that the young Black activists who had been present in their data in 2014 had disappeared eighteen months later (Freelon, McIlwain, and Clark 2016). Using the usernames of their young Black activist community as seeds, we found similarly that these users had largely disappeared-save for the few who themselves became famous, and joined the Black celebrity community instead.

In her influential theorization of *missing data sets*, the artist and educator Mimi Onuoha explains how the gaps in datasets provide "cultural and colloquial hints of what is deemed important," and further observes how "this lack of data typically correlates with issues affecting those who are most vulnerable in that context" (Onuoha 2018). In our #BlackLivesMatter dataset and the findings that it has enabled, these young Black activists constitute this precise

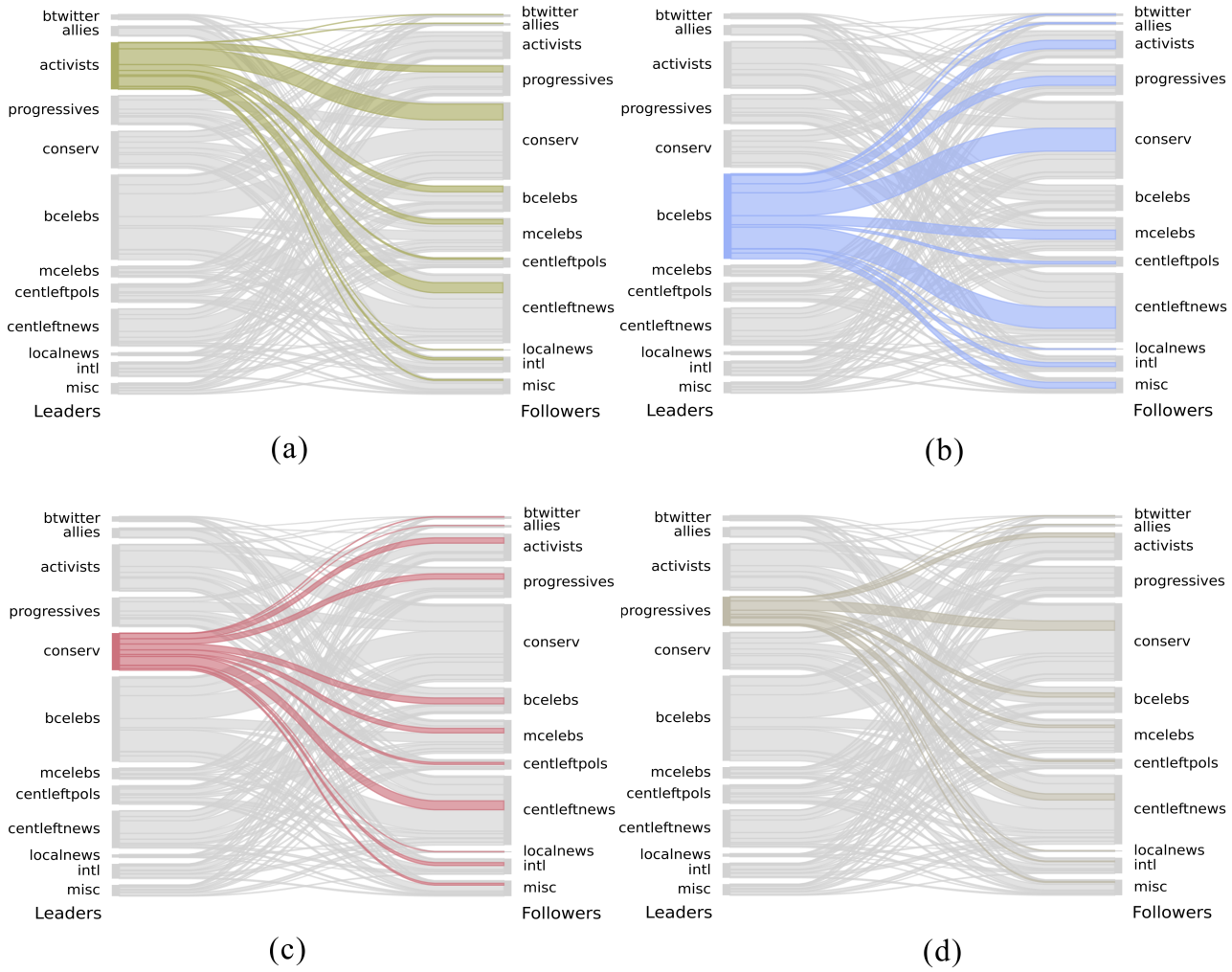

Figure 5: **Linguistic leadership**: The linguistic leadership of key communities including (a) activists, (b) Black celebrities, (c) conservatives, and (d) progressives with their contributions highlighted. The height of each bar is proportional to the overall number of words a community leads (left) or follows (right). The height of the connecting band is proportional to the number of words the community on the left leads the one on right.

form of missing data. Thus as we celebrate our findings for how they affirm the role of the Black Lives Matter movement in shaping a larger public conversation, we must also hold space for what our study cannot confirm, which is the equally crucial role played by the activists whose words are missing from our dataset. Even if their words are unrecorded, they are not unremembered.

Conclusion

In this project we describe a method of modeling semantic leadership across a set of communities associated with the Black Lives Matter movement, which has been informed by qualitative research on the structure of social media and of Black Twitter in particular. We describe our bespoke approaches to time-binning, community clustering, and connecting communities over time, as well as our adaptation of state-of-the-art approaches to semantic change detection and semantic leadership induction. We find substantial evidence of the leadership role of BLM activists and progressives, as well as Black celebrities. We also find evidence of the sustained engagement of the conservative community with this discourse, suggesting an alternative explanation for how we have arrived at the present political moment, in which "anti-woke" and "anti-DEI" policy is being enacted nation-wide. Contrary to the dominant narrative of conservative backlash against a movement that became too radical for the center to support, we find that the conservative community had been engaging directly with the BLM movement from its very start. The conservative community is the largest follower of word changes across the entire timespan of our study, and introduces many word changes as well, albeit few original conceptual keywords.

From our present vantage-point, in Spring 2025, we can now see the result of this sustained conservative engage-

ment: a distortion of the meaning of the words that Black Lives Matter worked so hard to bring to public attention, ultimately weaponizing the most powerful of these words to lead to policies that dismantle, rather than more fully realize, the movement's original goals. With this in mind, just as we recognize the role of the young Black activists who are missing from our data, we also recognize and hold space for the valuable work of all those involved in the Black Lives Matter movement. While the immediate gains of their work are currently being dismantled, the ideas behind their words remain a powerful force. We submit this paper as evidence of what words can accomplish when coupled with a commitment to action and an unwavering belief in the importance of equality and justice for all.

Acknowledgments

Thank you to Tanvi Sharma for her design work on the final visualizations. This project would not have been possible without the data collection work of Alex Fan. It also greatly benefited from discussions with André Brock and Jacob Eisenstein. This project has been supported by a grant from the Mellon Foundation (G-2211-14240) and by Emory's Department of Quantitative Theory & Methods.

References

Anderson, M. 2016. The hashtag #BlackLivesMatter emerges: Social activism on Twitter. Technical report, Pew Research Center.

Arif, A.; Stewart, L. G.; and Starbird, K. 2018. Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse. In *Proceedings of the ACM on Human-Computer Interaction*, CSCW, 1–27.

Bamman, D.; Dyer, C.; and Smith, N. A. 2014. Distributed representations of geographically situated language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 828–834.

Blevins, J. L.; Lee, J. J.; McCabe, E. E.; and Edgerton, E. 2019. Tweeting for social justice in #Ferguson: Affective discourse in Twitter hashtags. *New Media & Society*, 21(7): 1636–1653.

Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10008.

Bozarth, L.; and Budak, C. 2017. Is Slacktivism Underrated? Measuring the Value of Slacktivists for Online Social Movements. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1): 484–487.

Brock, A. J. 2020. *Distributed Blackness: African American Cybercultures*. NYU Press.

Brown, M.; Ray, R.; Summers, E.; and Fraistat, N. 2017. #SayHerName: A case study of intersectional social media activism. *Ethnic and Racial Studies*, 40(11): 1831–1846.

Bruckman, A. 2002. "Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet". *Ethics and Information Technology*, 4.

Card, D. 2023. Substitution-based Semantic Change Detection using Contextual Embeddings. In *The 61st Annual Meeting of the Association for Computational Linguistics*.

Choudhury, M. D.; Jhaver, S.; Sugar, B.; and Weber, I. 2016. Social Media Participation in an Activist Movement for Racial Equality. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10, 92–101.

Clark, M. D. 2024. *"We Tried to Tell Y'All: Black Twitter and the Rise of Digital Counternarratives"*. Oxford University Press.

del Rio, G. M. N. 2020. Oluwatoyin Salau, Missing Black Lives Matter Activist, Is Found Dead. https://www.nytimes.com/2020/06/15/us/oluwatoyin-salau-dead-aaron-glee.html. Accessed 2024-15-05.

Dunivin, Z. O.; Yan, H. Y.; Ince, J.; and Rojas, F. 2022. Black Lives Matter protests shift public discourse. *Proceedings of the National Academy of Sciences*, 119(10): e2117320119.

Dym, B.; and Fiesler, C. 2020. "Ethical and privacy considerations for research using online fandom data". *Transformative Works and Cultures*, 33.

Field, A.; Park, C. Y.; Theophilo, A.; Watson-Daniels, J.; and Tsvetkov, Y. 2022. An analysis of emotions and the prominence of positivity in #BlackLivesMatter tweets. *Proceedings of the National Academy of Sciences*, 119(35): e2205767119.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/. Accessed: 2024-15-05.

Freelon, D.; Lynch, M.; and Aday, S. 2015. Online Fragmentation in Wartime: A Longitudinal Analysis of Tweets about Syria, 2011–2013. *The ANNALS of the American Academy of Political and Social Science*, 659(1): 166–179.

Freelon, D.; McIlwain, C.; and Clark, M. 2018. Quantifying the power and consequences of social media protest. *New Media & Society*, 20(3): 990–1011.

Freelon, D.; McIlwain, C. D.; and Clark, M. 2016. Beyond the Hashtags: #Ferguson, #Blacklivesmatter, and the Online Struggle for Offline Justice.

Gallagher, R. J.; Reagan, A. J.; Danforth, C. M.; and Dodds, P. S. 2018. Divergent discourse between protests and counter-protests: #BlackLivesMatter and #AllLivesMatter. *PLOS ONE*, 13(4): e0195644.

Garg, N.; Schiebinger, L.; Jurafsky, D.; and Zou, J. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16): E3635–E3644.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Gillani, N.; and Levy, R. 2019. Simple dynamic word embeddings for mapping perceptions in the public sphere. In *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science*, 94–99.

Giorgi, S.; Guntuku, S. C.; Himelein-Wachowiak, M.; Kwarteng, A.; Hwang, S.; Rahman, M.; and Curtis, B. 2022. Twitter Corpus of the #BlackLivesMatter Movement and Counter Protests: 2013 to 2021. *Proceedings of the International AAAI Conference on Web and Social Media*, 16: 1228–1235.

Giulianelli, M.; Del Tredici, M.; and Fernández, R. 2020. Analysing Lexical Semantic Change with Contextualised Word Representations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 3960–3973. Online: Association for Computational Linguistics.

González-Bailón, S.; Borge-Holthoefer, J.; Rivero, A.; and Moreno, Y. 2011. The dynamics of protest recruitment through an online network. *Scientific reports*, 1(1): 1–7.

Grave, É.; Bojanowski, P.; Gupta, P.; Joulin, A.; and Mikolov, T. 2018. Learning Word Vectors for 157 Languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Hamilton, W. L.; Leskovec, J.; and Jurafsky, D. 2016a. Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, volume 2016, 2116.

Hamilton, W. L.; Leskovec, J.; and Jurafsky, D. 2016b. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, 1489–1501.

Jackson, S. J.; Bailey, M.; and Welles, B. F. 2020. *HashtagActivism: Networks of Race and Gender Justice*. MIT Press.

Jones, K.; Nurse, J. R.; and Li, S. 2022. Out of the Shadows: Analyzing Anonymous' Twitter Resurgence during the 2020 Black Lives Matter Protests. *Proceedings of the International AAAI Conference on Web and Social Media*, 16(1): 417–428.

Jules, B.; Summers, E.; and Mitchell, V. 2018. Ethical Considerations for Archiving Social Media Content Generated by Contemporary Social Movements: Challenges, Opportunities, and Recommendations. Technical report, Shift Collective.

Kim, Y.; Chiu, Y.-I.; Hanaki, K.; Hegde, D.; and Petrov, S. 2014. Temporal Analysis of Language through Neural Language Models. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, 61–65.

Klassen, S.; Kingsley, S.; McCall, K.; Weinberg, J.; and Fiesler, C. 2022. Black Lives, Green Books, and Blue Checks: Comparing the Content of the Negro Motorist Green Book to the Content on Black Twitter. *Proc. ACM Hum.-Comput. Interact.*, 6.

Kulkarni, V.; Al-Rfou, R.; Perozzi, B.; and Skiena, S. 2015. Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web*, 625–635.

Laicher, S.; Baldissin, G.; Castañeda, E.; Schlechtweg, D.; and im Walde, S. S. 2020. CL-IMS @ DIACR-Ita: Volente o Nolente: BERT does not outperform SGNS on Semantic Change Detection. arXiv:2011.07247.

Le-Khac, L.; Antoniak, M.; and So, R. J. 2023. #BLM Insurgent Discourse, White Structures of Feeling and the Fate of the 2020 "Racial Awakening". *New Literary History*, 53(4): 667–692.

Lui, M.; and Baldwin, T. 2012. langid. py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 system demonstrations*, 25–30.

Markham, A. 2012. FABRICATION AS ETHICAL PRACTICE. *Information, Communication & Society*, 15(3): 334–353.

Mendelsohn, J.; Vijan, M.; Card, D.; and Budak, C. 2024. Framing Social Movements on Social Media: Unpacking Diagnostic, Prognostic, and Motivational Strategies. *Journal of Quantitative Description: Digital Media*, 4.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.

Mundt, M.; Ross, K.; and Burnett, C. M. 2018. Scaling Social Movements Through Social Media: The Case of Black Lives Matter. *Social Media + Society*, 4(4): 2056305118807911.

Newman, M. E. 2006. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23): 8577–8582.

Newman, M. E.; and Girvan, M. 2004. Finding and evaluating community structure in networks. *Physical review E*, 69(2): 026113.

Onuoha, M. 2018. GitHub - MimiOnuoha/missing-datasets: An overview and exploration of the concept of missing datasets. — github.com. https://github.com/MimiOnuoha/missing-datasets. Accessed: 2024-15-05.

Peng, H.; Budak, C.; and Romero, D. M. 2019. Event-Driven Analysis of Crowd Dynamics in the Black Lives Matter Online Social Movement. In *The World Wide Web Conference*, WWW '19, 3137–3143. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-6674-8.

Press, A. 2019. A puzzling number of men tied to the Ferguson protests have since died. https://www.chicagotribune.com/2019/03/18/a-puzzling-number-of-men-tied-to-the-ferguson-protests-have-since-died/. Accessed: 2024-05-15.

Ray, R.; and Gibbons, A. 2021. "Why are states banning critical race theory?". https://www.brookings.edu/articles/why-are-states-banning-critical-race-theory/. [Accessed 15-05-2024].

Rudolph, M.; and Blei, D. 2018. Dynamic embeddings for language evolution. In *Proceedings of the 2018 World Wide Web Conference*, 1003–1011.

Soni, S.; Bamman, D.; and Eisenstein, J. 2022. Predicting Long-Term Citations from Short-Term Linguistic Influence. *Findings of the Association for Computational Linguistics: EMNLP 2022*.

Soni, S.; Klein, L.; and Eisenstein, J. 2021. Abolitionist Networks: Modeling Language Change in Nineteenth-Century Activist Newspapers. *Journal of Cultural Analytics*.

Soni, S.; Lerman, K.; and Eisenstein, J. 2020. Follow the Leader: Documents on the Leading Edge of Semantic Change Get More Citations. *Journal for the Association of Information Science and Technology*.

Spiro, E.; and Monroy-Hernández, A. 2016. Shifting Stakes: Understanding the Dynamic Roles of Individuals and Organizations in Social Media Protests. *PLoS One*, 11(10).

Stewart, L. G.; Arif, A.; Nied, A. C.; Spiro, E. S.; and Starbird, K. 2017. Drawing the Lines of Contention: Networked Frame Contests Within #BlackLivesMatter Discourse. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW): 96:1–96:23.

Tahmasebi, N.; Borin, L.; and Jatowt, A. 2019. Survey of Computational Approaches to Lexical Semantic Change. arXiv:1811.06278.

Theocharis, Y.; Lowe, W.; Van Deth, J. W.; and García-Albacete, G. 2015. Using Twitter to mobilize protest action: online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. *Information, Communication & Society*, 18(2): 202–220.

Tufekci, Z. 2013. "Not This One": Social Movements, the Attention Economy, and Microcelebrity Networked Activism. *American Behavioral Scientist*, 57(7): 848–870.

Tufekci, Z.; and Wilson, C. 2012. Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of communication*, 62(2): 363–379.

Walcott, R. 2024. #RIP Twitter: The Conditions of Black Social Media Platform Migration. [Accessed 15-05-2024].

Walsh, M. 2023. *Debates in the Digital Humanities 2023*, chapter The Challenges and Possibilities of Social Media Data: New Directions in Literary Studies and the Digital Humanities. Minneapolis: University of Minnesota Press.

Walsh, M. 2024. Recommendations for Using Social Media Data in Research—Whether You're in English or Info Science. Accessed 2024-06-09-2024.

Wijaya, D. T.; and Yeniterzi, R. 2011. Understanding semantic change of words over centuries. In *Proceedings of the 2011 international workshop on DETecting and Exploiting Cultural diversiTy on the social web*, 35–40. ACM.

Wilkins, D. J.; Livingstone, A. G.; and Levine, M. 2019. Whose tweets? The rhetorical functions of social media use in developing the Black Lives Matter movement. *British Journal of Social Psychology*, 58(4): 786–805.

Williams, R. 1985. *"Keywords: A Vocabulary of Culture and Society "*. Oxford University Press.

Xue, H. 2020. ""#BlackLivesMatter: The young Black activists using social media to lead the fight for equality". https://assembly.malala.org/stories/young-black-activists-to-follow-on-social. Accessed: 2024-15-05.

Yan, M.; Chiang, A. Y.; and Lin, Y.-R. 2024. From Posts to Pavement, or Vice Versa? The Dynamic Interplay between Online Activism and Offline Confrontations. *Proceedings of the International AAAI Conference on Web and Social Media*, 18(1): 1687–1701.

Yourish, K.; Daniel, A.; Datar, S.; White, I.; and Gamio, L. 2025. These Words Are Disappearing in the New Trump Administration. *The New York Times*. Accessed: 2025-03-07.

Paper Checklist

1. For most authors...

   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, and in fact this research seeks to directly intervene into and help to remedy existing socio-economic divides.

   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope?Yes, see the Abstract and Introduction.

   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made Yes, see the Background and Prior Work section, along with the various methods sections.

   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, see the Data and Implications and Limitations sections.

   (e) Did you describe the limitations of your work? Yes, see the Implications and Limitations sections.

   (f) Did you discuss any potential negative societal impacts of your work? Yes, see the Ethics and Privacy section, which discusses harms potentially brought about by social media research and our steps to mitigate them.

   (g) Did you discuss any potential misuse of your work? No, because we do not see any potential misuse of our work.

   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, see the Ethics and Privacy section, which discusses harms potentially brought about by identifying individual users and our steps to mitigate them. Should the paper be accepted, we will provide only derivative data so as to preserve anonymity. We will make our own code publicly available via GitHub.

   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes,and we take ethical considerations very seriously.

2. Additionally, if your study involves hypotheses testing...

   (a) Did you clearly state the assumptions underlying all theoretical results? NA

   (b) Have you provided justifications for all theoretical results? NA

(c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA

(d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA

(e) Did you address potential biases or limitations in your theoretical framework? NA

(f) Have you related your theoretical results to the existing literature in social science? NA

(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA

3. Additionally, if you are including theoretical proofs...

(a) Did you state the full set of assumptions of all theoretical results? NA

(b) Did you include complete proofs of all theoretical results? NA

4. Additionally, if you ran machine learning experiments...

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA

(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA

(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...

(a) If your work uses existing assets, did you cite the creators? Yes, see the Background and Prior Work and Data sections.

(b) Did you mention the license of the assets? NA

(c) Did you include any new assets in the supplemental material or as a URL? No, but we will put our code on GitHub if the paper is accepted.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes, see the Ethics and Privacy section, which discusses our decision to paraphrase rather than obtain consent from individual users; and our use of a "reasonably public" threshold for naming specific users, both employe due to concerns over exposure and harm.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes, see the Ethics and Privacy section and our response to 5d just above.

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see (FORCE11 2020))? NA

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see (Gebru et al. 2021))? NA

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...

(a) Did you include the full text of instructions given to participants and screenshots? NA

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA

(d) Did you discuss how data is stored, shared, and deidentified? NA

Appendix

Maxima	Additional Candidate Maxima
Sep 02, 2014	n/a
Dec 25, 2014	Dec 18, 2014
Aug 31, 2015	Mar 14, 2015; Apr 10, 2015; May 04, 2015
Jul 08, 2016	n/a
May 27, 2020	n/a
Aug 07, 2020	Sep 05, 2020; Sep 25, 2020

Table 3: **Maxima**: Additional candidate maxima that were merged into the final time bin segments.

Community	Short Label
Black Twitter	btwitter
Allies/Academics	allies
BLM Activists	activists
Progressives	progressives
Conservatives	conserv
Black Celebrities	bcelebs
Mixed Celebrities	mcelebs
Center/Left Politicians	centleftpols
Center/Left News Media	centleftnews
Local News	localnews
International Users	intl

Table 4: **Labeling Scheme**: Table translating between the long and short labels used throughout this paper.

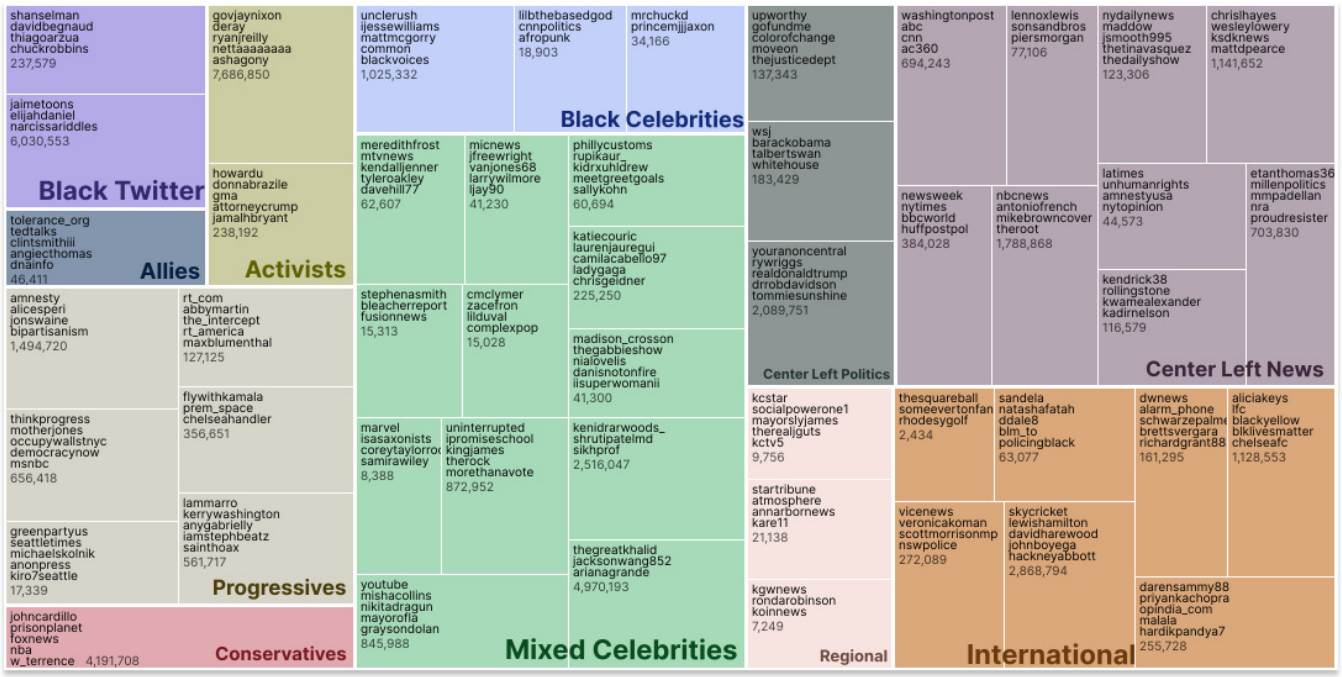

Figure 6: **Top Community Members**: The most central users of each subcommunity cluster, with color indicating final community grouping. Users with fewer than 3000 followers (as of September 2024) or with inactive accounts are not shown. Note that ranking by in degree leads to certain users (e.g. @realdonaldtrump) being included in communities in which they engage adversarially.

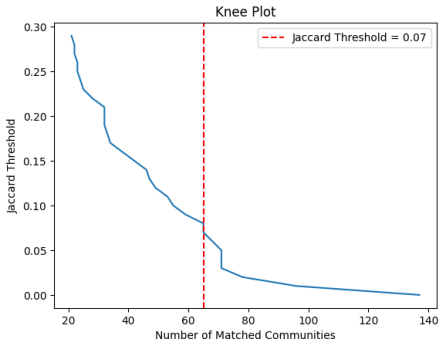

Figure 7: **Viable thresholds**: A knee plot showing the number of matched communities at Jaccard similarity thresholds between 0 and .3 in increments of .0125.