ABSTRACT

Title of Dissertation:	TALKING ABOUT JUSTICE: PREDICTING ACTOR ENGAGEMENT ON SOCIAL MEDIA AFTER A GALVANIZING EVENT
	Kimberly Ann Glasgow, Doctor of Philosophy, 2021
Dissertation directed by:	Associate Professor, Jessica Vitak, College of Information Studies

Social media contributes to discourse around and framing of major societal issues, and enables community formation, social change, and activism. It provides opportunities to engage in discourse, gain and share knowledge, and form ties with others around an issue, topic, or cause. This dissertation explores how justice, an important concept underlying social systems, is expressed in Twitter data in the context of high-salience, galvanizing local events, and leverages that information to predict whether newcomers to the issue will continue their digital engagement on the topic over time. It also attempts to quantify whether, and how much, a set of factors or dimensions previously associated with engagement in the physical realm contribute to digital engagement. These dimensions—*identity, emotion, effort, and social embeddedness*—are informed by prior work on social movements, digital activism, and related fields. Rather than rely on hashtags, this dissertation uses machine learning to detect justice-related Twitter activity. This advance in methods provides a richer understanding of discourse around a complex, multifaceted topic like justice. It allows deeper insight into the social media activity of newcomers to the justice community, and the networks they are embedded in. The approach is developed and applied first to Twitter data from Baltimore around the 2015 death of Freddie Gray from injuries sustained while in police custody, and the protests and riots that followed in Baltimore. To test for generalizability, the same approach is then applied to a second dataset, collected from Cleveland at the time of the death of Tamir Rice, who was shot and killed by police in 2014. Findings show that digital engagement in justice discourse on social media can be predicted, based on aspects of social embeddedness, emotion, and effort. To the degree that committed individuals are at the heart of social movements and efforts to spur social and civic change, and forming and being embedded in appropriate network structures is critical for channeling commitment into action and eventual success, this work contributes to greater understanding of these phenomena. Findings from this research could contribute to the design of technology to support civic engagement through social media platforms.

TALKING ABOUT JUSTICE: PREDICTING ACTOR ENGAGEMENT ON SOCIAL MEDIA AFTER A GALVANIZING EVENT

by

Kimberly Ann Glasgow

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Associate Professor Jessica Vitak, Chair
Assistant Professor Yla Tausczik
Associate Professor, Vanessa Frias-Martinez
Professor Jennifer Golbeck
Professor Philip Resnik, Dean's Representative

Dedication

To my husband, Ronald Schouten, in appreciation of his unwavering support.

Acknowledgements

I would like to acknowledge the tireless support of my advisors, Jessica Vitak and Yla Tausczik, throughout this process. My colleagues and mentors at the Johns Hopkins University Applied Physics Laboratory have set an example of professionalism and dedication to scholarship, and provided technical resources that enabled this research. I would also like to thank the members of my dissertation committee and the faculty of the iSchool that I have had the opportunity to work with and learn from.

Table of Contents

Dedication ii

Acknowledgements iii

Table of Contents iv

List of Tables vii

List of Figures ix

Chapter 1: Introduction 1

1.1 Predicting Justice Engagement 5
1.1.1 Identity 6
1.1.2 Effort 7
1.1.3 Emotion 8
1.1.4 Social Embeddedness 8
1.2 Context 10
1.3 Background on Freddie Gray incident 12
1.4 Background on Tamir Rice incident 14
1.5 Contributions 15
Dissertation Structure 16
Chapter 2: Related Work 18

Justice 18
Social movements 20
Social media 23
2.4 Social media and justice, policing, and political movements 25
2.5 Social networks, justice perceptions, and collective action 30
2.6 Dissertation Research Questions 34
Chapter 3: Approach and Methods 36

3.1 Overview of Approach 36
3.2 Twitter data from Baltimore and Cleveland 38
3.3 Processing tweet text and emoji 43
3.3.1 Text processing 43
3.3.2 Empath 44
3.3.3 Emoji 45
3.4 Image classes and classification 48
3.5 Hashtags and hashtag categories 51
3.5.1 Baltimore hashtags and hashtag categories 52
3.5.2 Cleveland hashtags and hashtag categories 55
3.6 Tweet classification for justice 58

3.6.1 Developing a ground truth dataset for tweet classification........................ 58
3.6.2 Tweet classifiers for Baltimore and Cleveland... 62
3.7 Simpler models for tweet classification for justice.. 69
3.8 Distinguishing and characterizing highly engaged vs. cursory actors in justice discourse 73
3.9 Determining actor-level indicators for predicting sustained justice engagement 77
3.9.1 Social network analysis of tweet-based justice networks.......................... 77
3.9.2 Empath variables for predictive modeling... 80
3.9.3 Creating features for predictive models of user engagement..................... 81
3.10 Summary of approach and methods... 83
Chapter 4: Analysis and findings.. 85

4.1 Overview of engagement models... 87
4.2 Identity 92
4.3 Emotion.. 93
4.4 Effort 94
4.5 Social Embeddedness... 94
4.6 Topical and General... 97
4.7 Review of engagement models at different thresholds 97
4.8 Summary 98
Chapter 5: Discussion 99

5.1 Historical and event-driven perspective on Cleveland and Baltimore.............. 99
5.2 The role of identity variables in predicting continued justice engagement 104
5.3 The role of effort variables in predicting continued justice engagement........ 106
5.4 The role of emotion variables in predicting continued justice engagement ... 107
5.5 The role of topical variables in predicting continued justice engagement...... 109
5.6 The role of social embeddedness in predicting continued justice engagement 111
5.7 Social embeddedness and justice tweeting 112
5.8 Limitations 113
Chapter 6: Conclusion.. 115

6.1 Contributions of this study.. 116
6.2 Future research and recommendations.. 122
Appendix A: Emoji classes 124

Appendix B: Coding Guidance for tweet classification 128

Appendix C: Empath.. 129

Appendix D: Model comparisons at different thresholds 134

Appendix E: Hashtags identified for Baltimore and Cleveland..................... 141

Appendix F: Correlations for Baltimore and Cleveland variables used in logistic regression 146

Glossary 147

Bibliography 148

List of Tables

Table 1: Positive and Negative emotions for Emoji annotation, drawn from Goleman (1995) 46
Table 2: Inter-annotator agreement for emotion classification of emoji. Two annotators assessed all emoji occurring at least 200 times in the Baltimore dataset for emotion, judging the positive or negative emotional valence of the emoji. 47
Table 3: Inter-annotator agreement for emoji classification for person class. Two annotators assessed all emoji occurring at least 200 times in the Baltimore dataset for representing a person........ 47
Table 4: Image classes and class descriptions. This guidance was used to assess images for use in creating Deep Learning image classifiers........ 49
Table 5: Frequently occurring hashtags in the Baltimore tweet corpus. Hashtags pertaining to Freddie Gray, the rioting that took place after his death, or justice are bolded........ 52
Table 6: Top relevant hashtags for Freddie Gray, Riot, and Justice. Of the thousand most frequently occurring hashtags in the data, a few dozen were relevant to justice, Freddie Gray, or the riots. Justice-related hashtags extended beyond calls for justice for Freddie Gray, evoking other victims of police violence, and larger social movements and actions........ 54
Table 7: Frequently occurring hashtags in the Cleveland tweet corpus. Several pertain to justice or Tamir Rice. Most are not related to the topic. 56
Table 8: Top relevant hashtags for Tamir Rice and Justice. Of the thousand most frequently occurring hashtags in the data, a few dozen were relevant to the topic. Justice-related hashtags extended beyond calls for justice for Tamir Rice, evoking other victims of police violence, and larger social movements and actions. Several refer to counter-narratives or movements supporting police, and downplaying systemic injustice and violence in the criminal justice system. Fewer relevant hashtags are identified in this dataset than were identified for Baltimore........ 57
Table 9: Justice class annotation guidance for supervised classification for tweets. Example tweets that meet or do not meet the criteria for justice are provided... 59
Table 10: Annotator agreement for justice annotations........ 62
Table 11: Confusion matrices for specific machine learning models on Baltimore justice data (binary classification of justice/not justice, n=2462, 80/20 split).... 63
Table 12: Metrics of model performance for specific machine learning models on Baltimore justice data. 63
Table 13: Performance of ensemble model for Baltimore justice classification data (binary classification of justice/not justice, n=2462)........ 65
Table 14: Confusion matrices for machine learning models for Cleveland justice data (binary classification of justice/not justice, n=2271, 80/20 split)........ 65
Table 15: Metrics of model performance for specific machine learning models on Cleveland justice data. 66

Table 16: Performance of ensemble model for Cleveland justice classification (binary classification of justice/not justice, n=2271). 68
Table 17: Confusion matrices for best performing machine learning models using hashtag or hashtag and language features on Baltimore justice data (binary classification of justice/not justice, n=2462, 70/30 split) 70
Table 18: Metrics of model performance for machine learning models using hashtag, language, or hashtag and language features on Baltimore test data 70
Table 19: Confusion matrices for best machine learning models using hashtag, language, or hashtag and language features on Cleveland justice data (binary classification of justice/not justice, n=2271, 70/30 split): 72
Table 20: Metrics of model performance for machine learning models using hashtag or hashtag and language features on Cleveland test data.................................... 72
Table 21: Summary of dimensions and associated variables. These variables will be measured at T2, the time period of the riots, for users who tweeted about justice, and should help predict if an actor will remain engaged in justice discourse at T3, after the riots ended and when general interest and news media coverage has ebbed. 76
Table 22: Tie types used in social network analysis.. 78
Table 23: Sample social network metrics considered for assessing social embeddedness of actors. These metrics helps characterize the prominence, or centrality, of actors as well as whether they are situated within the main component or a smaller component of the network. 79
Table 24: Empath variables retained for predictive models 80
Table 25: Variables used in predicative models for user engagement after a galvanizing event. These variables were included in logistic regression models for predicting sustained engagement with justice discourse in T3, post event. For each variable, the aggregation method, and relevant dimensions are shown. 81
Table 26: Logistic regression model for sustained justice engagement in Baltimore, at a threshold of engagement of 4 or more tweets during T3. This model achieves a Pseudo R-Square (optimistic) of 0.643. (Sig. codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1)... 87
Table 27: Logistic regression model for sustained justice engagement in Cleveland, at a threshold of engagement of 4 or more tweets during T3. This model achieves a Pseudo R-Square (optimistic) 0.719. (Sig. codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1).. 89

List of Figures

Figure 1: Overview of tweet processing and classification approach to support analysis. Tweets are ingested and their associated images are retrieved. The tweet text undergoes several processing steps to improve its usefulness for classification. Additional enrichment features are computed. All of these features are available to be used in the classification step. After classification, a dataset of relevant, enriched tweets is produced for further analysis and the predictive model of user engagement in justice discourse.. 39

Figure 2: Tweets per day in Baltimore. From an average of about 100,000 tweets per day, activity spikes around the time of Freddy Gray's death and the violent protests in Baltimore. 41

Figure 3: Tweets per day in Cleveland. While there is an increase in activity around the time of Tamir Rice's death, the spike is much less pronounced than was observed in Baltimore. 43

Figure 4: Image classifier performance for all image classes. A Receiver Operating Curve (ROC) is shown. 51

Figure 5: Example tweet containing an image that conveys information for justice classification 61

Figure 6: Overlap in agreement across Baltimore justice classifiers. Most tweets are classified positive by multiple classifiers. 64

Figure 7: Overlap in agreement across Cleveland justice classifiers. With the exception of the normalized text classifier, most tweets are classified positive by the remaining classifiers. 67

Figure 8: Justice tweets per day in Baltimore. From a lower baseline rate (T1), justice tweeting increases substantially in T2, then drops in T3, though activity remained elevated compared to the earlier T1 rate. 85

Figure 9: Justice tweeting activity in Cleveland. From a lower baseline at T1, activity spikes dramatically during the peak of the event (T2). A smaller increase persists through T3, the post-event time period. 86

Figure 10a: Receiver Operating Curve (ROC) for Baltimore model. The area under the curve is slightly higher for Cleveland than Baltimore. 91

Figure 11b: Receiver Operating Curve (ROC) for Cleveland model. The area under the curve is slightly higher for 92

Figure 12: Network of Baltimore justice actors in T2. This network includes over 25,000 actors, with most in the main component. Actors who remain engaged in T3 are both colored red and highlighted, to improve their visibility. Nearly 90% of these engaged actors are in the main component. The main component is visible in the center, while smaller components of connected actors are shown to the right, and isolated actors are displayed on the left. Actors are connected to other actors if they mentioned, replied to, or retweeted them. 95

Figure 13: Network of Cleveland justice actors at T2. This network includes over 27k actors, most of whom are in the main component, as was observed in Baltimore. Actors who remain engaged in T3 are both colored red and highlighted, to improve their visibility. The main component is visible in the center, while smaller components of connected actors are shown to the right, and isolated actors are displayed on the left.. 96

Figure 14: Tweet from the Cleveland police department during a pro-police rally in the city....... 101
Figure 15: Justice tweet from Cleveland that referenced the Empath concept "weapon"....... 110

Chapter 1: Introduction

The potential impact and power of social media to organize, support, sustain, and ignite societal change, social protest, and civil unrest helping create a shared language, purpose, and sense of identity among members (Conover et al., 2013). is undeniable. Platforms like Twitter can serve as a mechanism for allowing individuals to both become informed about and spread information on events and issues relating to social movements (Gleason, 2013; Lotan et al., 2011), to share thoughts, opinions or emotions, to promote civic engagement and action (Gordon et al., 2013), or to form ad hoc discursive communities around aspects of the public sphere (Bruns & Burgess, 2011).

The last decade provides numerous examples to highlight the power of Twitter and other social media in evoking change Activity on social media played a role in the collective framing processes of social movements such as the Occupy Wall Street movement in 2011 (#Occupy),Likewise, during the Arab Spring revolution, which began in late 2010 and continued well into 2011 in several countries in Northern Africa and the Middle East, researchers identified spikes in conversations about liberty, democracy, and revolution on blogs and Twitter immediately before the largest protests (Howard et al., 2011). Work on social upheaval in Turkey found that both the roles of individuals and amount of individual influence on social media evolved during the course of events (Varol et al., 2014). Social media activity has even been demonstrated to be predictive of participation in future protests on the ground (De Choudhury et al., 2016). Such work lends credence to an ongoing intertwining of the online world of social media and the offline, physical world.

Rather than focus on a specific activist movement, such as #BlackLivesMatter or #OccupyWallStreet, in this dissertation I examine a more foundational concept – the concept of justice itself, as expressed through social media discourse. Perceptions of justice are fundamental to citizen compliance with the law and help underpin social order and safe communities (Schulhofer et al., 2011). I explore and advance understanding of how social media-enabled communities concerned with issues of justice emerge and develop in the aftermath of highly salient local events. Perceptions of justice are formulated in social settings and dependent on social information (Cropanzano, Ambrose, Masterson, et al., 2015). I examine the individuals involved in justice-related social media discourse within two physical communities (Baltimore, MD, and Cleveland, OH), as well as the specific networks of participation they are embedded in. The potential for networks to enable or impede the diffusion, reinterpretation, and amplification of justice perceptions has previously been proposed (Shapiro et al., 2008); a goal of this dissertation is to shed empirical light on this topic.

This research focuses attention on specific geographic regions (cities) that have experienced a galvanizing justice-related event. Prior work considering social media behavior from a spatiotemporal perspective in the context of protests against perceived injustice in Ferguson, MO after the killing of Michael Brown in 2014 has found events show distinctive linguistic and temporal patterns locally as compared to globally (He et al., 2015). Geotagged tweets from St. Louis peaked earlier, consistent with unfolding events in Ferguson, while other U.S. cities peaked later, after then-President Obama addressed the nation. Tweets from the St. Louis area also contained more detailed or specific location and personality language, such as mentions of the street where the shooting took place (Florissant) and Governor Nixon.

This work also complements prior work using hashtags to define relevant data (Bonilla & Rosa, 2015; Bruns & Burgess, 2011; Gleason, 2013; Raynauld et al., 2018). It proposes methodological improvements to dataset gathering to help address the concern that big data analysis of social media is over-reliant on hashtags to identify relevant postings, which can introduce bias into the data. For example, highly engaged users are likely to drop popular hashtags over time as redundant, many hashtags fail to reach thresholds of popularity sufficient to be included, and users may discuss an event or issue without resorting to hashtags (Tufekci, 2014). To move beyond this limitation, the present study develops and implements machine learning models to identify tweets related to justice.

Data collected and analyzed in this dissertation captures the time period when events unfolded around the deaths of two young African Americans. The first dataset includes 6 million tweets surrounding the death of Freddie Gray, in Baltimore, MD in April 2015, including his arrest and death and the ensuing protests, riots, and initial criminal justice system activity relating to the officers involved in the case. The second dataset, which was used to demonstrate generalizability with the approach applied to the Baltimore data, includes 8 million tweets surrounding the death of 12-year-old Tamir Rice, who was shot by police in 2014 in Cleveland, OH.

This work extends prior work on patterns of temporal activity and user participation around galvanizing events (Conover et al., 2013; De Choudhury et al., 2016). It examines justice-related discourse across a timeframe that encompasses a galvanizing event, organizing the timeframe into a smaller set of three event-driven phases. The first phase (T1) precedes Freddie Gray's arrest, and ends on the day of his death. It provides a baseline for justice related social media discourse participants and networks in Baltimore. The second phase (T2) includes protests, Freddie Gray's funeral, and the riots. The third phase (T3) covers post-riot activity, including the filing of charges against the officers involved in Freddie Gray's arrest. Such a division of the data allows differentiation between individuals who were active in justice discourse prior to the galvanizing event and those who only began participating *during* it. A similar separation of the data was applied to Cleveland dataset, with a baseline preceding the shooting of Tamir Rice (T1). The second phase encompasses Rice's death and the most intense protests in Cleveland (T2), followed by the third phase (T3).

This dissertation examines characteristics of individuals who maintain interest in justice after the galvanizing event is over—when public attention wanes—and builds models to predict who stays engaged with justice at later phases of the timeframe, and how they differ from those whose participation was ephemeral. The approach considers their social media behaviors and postings, as well as changes in identity presentation, affect, or network position within the justice network. Prior work observed social media discourse became less concentrated over time by looking the small set of users who were active early (Varol et al., 2014). I examine whether this pattern holds for justice-related discourse and participants in the Baltimore and Cleveland communities.

This study focuses on community discourse relating to justice, rather than selecting specific activists or organizations to focus on, which provides a broader community-level perspective on the issue. It applies machine learning methods to assess relevance of tweets, going beyond tweet text to incorporate meaning conveyed by images, emoji, and other features of the tweet. It extends observations about temporal patterns of social media activity around powerful events, and about "cursory" participants as compared to those who are more deeply engaged with an issue. It evaluates the extent to which behavioral, content-based, identity, emotion, and network factors are related to ongoing level of engagement after a salient event, and can be used in predictive models of user engagement.

In the following sections, I provide a brief overview of the primary factors used in building the models predicting continued engagement in justice discourse, as well as a summary of the two galvanizing events. This chapter concludes with more details on the contributions of the dissertation.

1.1 Predicting Justice Engagement

This dissertation helps bridge the gap between what has been learned through study of social movements, civic action, and justice in the physical world with how these phenomena transpire in the digital realm. By applying a computational social science approach, we can gain insight into how digital traces connected to one's identity, emotion, effort, and social embeddedness contribute to ongoing justice engagement at a much larger scale (i.e., across the population of a city active on a social media platform) and with finer temporal granularity than would be realistic with more traditional social science methods such as surveys or ethnographic research.

1.1.1 Identity

Presentation of personal identity is how one represents one's self in terms of personal attributes, affiliations, causes, and activities (Goffman & others, 1959). This identity is the public face of who we are, the face we show to others. Our personal identity can be linked to many social identities, or cognitive interpretations of our membership in one or more groups (Turner et al., 1979), that may be evoked in salient contexts. A social movement or other group will typically develop a collective identity, a set of cognitions shared by its members (Taylor et al., 1995).

Recent theorizing on how social media is transforming civic engagement, political participation, and social movements suggests that the primacy of collective identity in mobilizing members and motivating action is slipping in this new environment. It is being complemented, if not supplanted, by a personalization of participation in political and civic activities and social movements. Connective action (Bennett & Segerberg, 2012), digitally networked participation (Theocharis, 2015), and individualized collective action (Bennett & Segerberg, 2011) are contrasted to a monolithic, proscribed, collective approach to social and political change.

Prior work has established a distinction between "cursory" and more deeply engaged participants in social media discourse who continue to communicate about the issue over time, and has found that *identity* is one aspect of this distinction. That work found that the communicative behavior of those who stayed engaged reflected increased collective identity, as measured by increases in Linguistic Inquiry and

Word Count (LIWC) variables (Tausczik & Pennebaker, 2009), such as first-person plural (forms of "we"). Other work (Varol et al., 2014) observed some active participants had changed screen names as "signs of protest." In light of this, I anticipate that more engaged participants at T3 in my datasets will be more likely to have changed some aspect of their presentation of identity on Twitter, as reflected by their user profile description, screen name, or profile image at T2, or to have adopted language more reflective of a personal identity linked with the issue of justice.

For this research, I focus on individual actors involved in justice discourse, and consider personal identity as it might be related to the issue of justice. I do not attempt to determine whether that individual has joined a specific social movement, such as Black Lives Matter, or adopted its collective identity. In this research, I measure identity/identification through both Twitter profile-based variables (e.g., changes to screen name, profile image or description) and content-based ones.

1.1.2 Effort

Prior work has found social media users' topical diversity, breadth of information generation, and emotional intensity tends to exceed that observed from news media (Hong et al., 2016). I expect that the language and content shared by engaged users may incorporate elements that reflect a richer or more nuanced view of the topic. These users may employ a more diverse set of topical hashtags, or incorporate other features such as images, that will be less commonly used by cursory participants. For example, rather than solely commenting on justice in conjunction with references to Freddie Gray or riots—high profile, salient aspects of events in

Baltimore at the time—engaged users may have justice tweets that are not simultaneously Freddie Gray or riot-related. Individuals who remain engaged at T3 are thus likely to communicate more richly about justice at T2 than cursory participants. Thus, increased frequency of incorporation of indicators of effort, such as images of protests or police in their justice tweets, or use of relevant, evocative hashtags, should be predictive of continued engagement in justice discourse.

1.1.3 Emotion

The importance of emotional investment or connection as a component of activism or involvement in social or political change has long been noted (Klandermans, 1996). Individuals who are engaging in justice-related discourse with their local communities might experience a range of emotions, from the negative such and anger or sadness over perceived injustice or brutality, to pride and hope that a unified community might be able to bring about change. They might express different emotions to different actors within the community, supporting and encouraging those with whom they agree, or disapproving or condemning those whose actions are not aligned with their positions. This work examines whether the presence of emotion in justice related discourse, rather than its absence, is predictive of ongoing engagement.

1.1.4 Social Embeddedness

The importance of social ties in mobilization has been well established in social movement theory (Diani, 2015). In this work, I use social network measures related to actor centrality and position to operationalize social embeddedness.

Social embeddedness captures the assertion that both individual and institutional behavior are so constrained by ongoing social relations that to construe them as independent is a grievous error (M. Granovetter, 1985). I anticipate that engaged actors will be more deeply *socially embedded* in the justice discourse community, as reflected by their positions in its social network. For example, they may connect with more members of the justice discourse community, whether by sharing their content via retweeting, by replying, or by mentioning them. Actors who engage with a larger number of others have been observed to have more affective commitment to the group (Lee & Kim, 2011) and show higher retention in organizational contexts (McPherson et al., 1992). Prior work has also indicated that perceptions of justice are socially shaped (Cropanzano, Ambrose, Masterson, et al., 2015) and that networks provide a mechanism for social transmission of justice information (Shapiro et al., 2008).

I expect that newly engaged members will not be isolates at T3 (i.e., individuals who remain unconnected to any other member of the network despite tweeting many times about justice), but will have connected with actors from T1 and with each other, increasing their opportunities for information flow and increasing their social capital (Borgatti et al., 2009). Engaged actors will be likely to hold a more central network position than cursory actors. Moving from the periphery to a more central position as one becomes more integrated within a community has been observed in social network contexts. More central positions in networks that share information can correspond with being generally influential or being focal points for information flow (Freeman, 1979). A tendency to form ties with actors who are already highly connected, or "preferential attachment," has been well documented in social networks (Newman, 2001).

In contrast to previous findings, in which early adopters appeared to become less active and eventually drift away after a galvanizing event (Conover et al., 2013), I do not foresee a large proportion of early justice participants in our localized contexts losing interest in this topic. Conover examined social media participants in the Occupy movement at a global level, and used *#occupy*-related hashtags to identify relevant tweets. I anticipate that interest in the complex and evolving issue of justice in Baltimore and Cleveland may be sustained by individuals in their respective cities, and that a machine-learning based method for identifying relevant data will improve detection over a hashtag-driven approach. Further, local individuals are more likely to possess unique, relevant information in exceptional circumstances, and may be highly motivated to share with others (Sutton et al., 2008).

Prior work on activist movements has found that the number of new participants peaked during important events, as does the raw amount of activity (De Choudhury et al., 2016). Only a subset of those actors continued to discuss the topic after the peak has passed (Olteanu et al., 2015). While network analysis was outside the scope of that work, it may be suggestive of patterns in my datasets. Peak network size should occur during the salient event, with post-event network size diminishing, though still greater than during the baseline established pre-event.

1.2 Context

Both high-salience justice events examined in this research are embedded in a broader social context. These incidents occurred at a time when the United States had experienced a number of high-profile police killings of young black men and boys. Concerns about systemic bias in policing and the criminal justice system, and about racial disparities in police violence, had attracted renewed national attention and galvanized social action, both online on social media platforms and in organized protests and campaigns on the ground.

The hashtag #BlackLivesMatter was coined in the aftermath of the acquittal in 2013 of the man who stood trial for the murder of Trayvon Martin, a black teen who was shot to death in Florida. Later, this hashtag helped spark the eponymous Black Lives Matter activist movement (Bonilla & Rosa, 2015), which campaigns against police brutality and systemic racism.

Social science and criminal justice research, as well as investigations by the Department of Justice, provided evidence that these concerns are well-founded. Significant disparities exist in types of behaviors that are criminalized, enforcement, arrest rates, prosecution rates, conviction rates, sentencing, and parole with respect to race and class in the United States (Reiman & Leighton, 2015; Tonry, 2010). While President Lyndon Johnson's Commission on Law Enforcement and Administration of Justice proposed objectives to face the challenge of crime, including confronting sources of crime (poverty, lack of education, substance abuse), improving professional standards and training, and eliminating injustice throughout the system in 1967 (Douglas, 1967), actions to achieve these objectives have been uneven. The Department of Justice has undertaken nearly 70 formal investigations into whether "a pattern or practice of conduct by law enforcement officers" that violates Constitutional or federal rights exists in a police department (*Justice Department*

Releases Report on Civil Rights Division's Pattern and Practice Police Reform Work, 2017). Both the Baltimore Police Department and the Cleveland Division of Police have been found to engage in such practices, including:

- making unconstitutional stops, searches, and arrests (Baltimore);
- using enforcement strategies that produce severe and unjustified disparities in the rates of stops, searches and arrests of African Americans (Baltimore);
- using excessive force (Baltimore and Cleveland); and
- retaliating against people engaging in constitutionally-protected expression (Baltimore).

These patterns arose through systemic deficiencies in policies, training, supervision, and accountability. As a consequence, police failed to perform effectively and within the bounds of the federal law, and communities suffered. Both Baltimore and Cleveland have entered into consent decrees (Cleveland for the second time) to help remediate this situation.

1.3 Background on Freddie Gray incident

The first event considered in this dissertation is the arrest and death of Freddie Gray, a young African-American man, and the ensuing protests, riots, and initial criminal justice system activity relating to the officers involved in Mr. Gray's arrest. On April 12, 2015, Mr. Gray was arrested in the Sandtown-Winchester neighborhood of Baltimore, MD, after making eye contact with a police officer and fleeing. He was placed in the back of a police van to be transported to the nearby police station. At some point after his arrest and before arriving at the station 45 minutes later, he sustained severe injuries, including three fractured vertebrae and injuries to his larynx. His spine was 80% severed at the neck. He had fallen into a coma. Mr. Gray was transported to a shock trauma center; however, despite extensive surgery, Mr. Gray died from his spinal injuries one week later on April 19, 2015.

These events were widely reported in the news media. While local news outlets, such as the *Baltimore Sun*, began reporting on Mr. Gray by 18 April, the story did not garner national attention until days later. Local public response to the events included protests that began on April 18 and continued through the week. On April 25, violence erupted at a protest, and injuries and arrests were reported. On April 27, the day of Freddie Gray's funeral, violence escalated to rioting, looting, and arson. A state of emergency was declared in Baltimore, the Maryland National Guard was activated, and a curfew was declared. Hundreds of people were arrested. By April 29, cleanup of the extensive damage caused in the riots was well under way in the city, though many businesses were not yet able to reopen.

Considerable controversy raged in the wake of these events. The legality of the arrest, how Mr. Gray sustained his injuries, whether the police behaved criminally and caused his death, or negligently and failed to provide requested medical care, were debated throughout the city. Others felt the police were being unfairly judged. Broader issues of injustice, discriminatory policing, and brutality were raised, and the U.S. Department of Justice announced an investigation into the Baltimore Police Department. Criminal charges were announced against the six officers involved in the arrest and transport of Freddie Gray on May 1, 2015, after the medical examiner's office ruled that Mr. Gray's death was a homicide. On May 21, 2015, a grand jury returned indictments against the officers. During these events, government officials and the Baltimore Police Department used Twitter to communicate with the city, as shown in this tweet from the official police department account:

> *A group of protesters remain in front of the Western District.*
>
> *#PeacefulProtest.*

Throughout this time, local Baltimore citizens and journalists also used Twitter to provide updates and insights, or to share their views or opinions.

1.4 Background on Tamir Rice incident

On November 22, 2014, Tamir Rice, a 12-year-old African-American boy, was shot in in a park in by a 26-year-old white police officer named Timothy Loehmann. Rice was carrying a toy gun. He was shot within seconds of the police arriving on the scene, and died from his injuries at the hospital overnight.

In the wake of the shooting, protests and public outcry broke out in Cleveland, although they were relatively minor. However, on November 25, 2014, the day after a grand jury decision to not indict the police officer who fatally shot Michael Brown, the Cleveland protests became more prominent. That day, about 200 protesters marched from Public Square to the Cleveland Memorial, causing the latter to be shut down temporarily. Rice's family pleaded with the protesters to remain peaceful in their activities, saying, "Again, we ask for the community to remain calm. Please protest peacefully and responsibly." On November 26, surveillance video of the incident was released to the public. Unlike Baltimore, Cleveland did not experience rioting after the killing of Tamir Rice, and the National Guard was not deployed.

A funeral service for Tamir Rice was held on December 3, 2014, with approximately 250 people in attendance. On December 5, Ohio Governor John Kasich established a task force to address community-police relations in response to Rice's shooting and other similar incidents. Tamir Rice’s death was not declared a homicide until mid-December. The investigation was transferred to an outside jurisdiction in January 2015. Officer Loehmann, who killed Tamir Rice, was eventually fired in 2017 for lying on his employment application. Neither officer involved in the incident was ever criminally charged. However, the Rice family did receive a $6 million settlement for the wrongful death of their son from the city of Cleveland.

Rice’s death has been cited as one of several police killings in 2013 and 2014 that ‘sparked’ the nationwide Black Lives Matter movement.

1.5 Contributions

This study makes several contributions and extensions to literature on social movements and social media, and it offers methods recommendations for future event-based studies of social media data. It is grounded in analyzing a large corpus of social media (Twitter) activity and events at the local (city) level, as experienced and reflected by the community going through a devastating event that, for many, crystallized ongoing issues of justice and policing within their respective cities.

A primary contribution of this research arises through the development and demonstration of an approach used to predict ongoing social media engagement around the issue of justice. My approach is informed by work in the social sciences that helped explain participation in social movements, activism, civic engagement, and related phenomena. I adapt the dimensions of identity, social embeddedness, emotion, and effort to digital engagement and social media. Taking this approach allows me to both predict ongoing justice engagement by Twitter users, and to quantify whether (and how much) each of these dimensions contributes to predicting ongoing engagement in social justice discourse.

Additionally, I demonstrate the generalizability of my findings from one high-salience, galvanizing local event—the death of Freddie Gray from injuries sustained in police custody, and the protest and riots that followed in Baltimore in the spring of 2015—to the shooting of Tamir Rice by police in Cleveland.

Further, I demonstrate the applicability of a computational social science approach to understanding digital engagement. This approach extends more traditional data gathering from surveys of a small set of actors, by using data derived from social media behaviors of participants from the community during an event. Understanding who is more likely to remain active and engaged with justice discussions can help researchers in creating new tools and technology to mobilize and galvanize political participation through social media platforms. It can inform socio-technical researchers interested in digital engagement relating to social movements, political and civic action, and networks and identity.

Dissertation Structure

This dissertation will proceed as follows. Chapter 2 presents a synthesis of related work on justice, social movements, and social media, concluding with the four research questions that guide analyses. Chapter 3 presents the methods for this study, focusing on how the two datasets of tweets were collected, processed and enriched, how machine learning models were used to identify justice-related tweets, and how that curated dataset of justice tweets was used to model and predict the future engagement of users in justice discourse. Chapter 4 presents findings from the models for the two galvanizing events in Baltimore and Cleveland. Chapter 5 explores these findings in more detail, connecting this work to prior research, and details the limitations of this research. Chapter 6 provides final thoughts on the impact of this research and similar research on the way social media can facilitate wider social change, and it provides suggestions for future research in this space.

Chapter 2: Related Work

Justice

The orientation of this research is toward the criminal justice system. This is the system involved in enforcing the criminal law, and deterring, investigating, prosecuting, and sanctioning legal violations that are designated as criminal (offenses to society), rather than civil (offenses to other individuals). However, some findings from the organizational justice literature may prove relevant. Organizational justice examines how employees perceive fairness in the workplace, and the impact of these perceptions on job performance, organizational commitment, and similar outcomes. Several major perspectives have been used in the social sciences to recognize, understand, or model justice in society.

Mechanisms for controlling behavior are integral to human societies, and systems of norms and laws to support these functions are ubiquitous (Sunshine & Tyler, 2003). Justice, and the criminal justice system, are primary levers for shaping behavior and meting out consequences in our society. How justice is understood, and why people follow the law, are enduring questions that have led to a number of models or perspectives on justice.

One fundamental insight into justice is that people are not solely *instrumental* in their perspectives about justice. That is, individuals do not prefer outcomes simply because they personally benefit from them, regardless of how those outcomes were achieved. Instead, people also weigh whether the procedures used by police or other authorities were fair, and were consistently and appropriately applied. *Procedural* *justice* contributes to perceptions of fairness and legitimacy in the community (Tyler, 1994). When it is present, people are more likely to feel heard, conform to the law, and cooperate with authorities (Schulhofer et al., 2011). *Relational* models of justice build upon this earlier justice work, layering social concerns into the psychology of justice perception. The relational perspective considers how social identity and status intersect with perceptions of justice. It holds that individuals interpret interactions with authorities such as police as informative of the relative status of their social group. Interactions in which police exhibit neutrality instead of bias, treat group members with dignity, and behave in a trustworthy fashion are generalized as validating the group's position in society, and help establish expectations for future interactions with authorities (Cropanzano, Ambrose, Blader, et al., 2015; Tyler, 1994), as well as influencing levels of cooperation and conforming with the law. While this research does not seek to extend these theories or models, these concepts inform the methodological approach.

A study on the views of local residents following the death of another young Black man, Michael Brown in Ferguson, Missouri, at the hands of police and the riots that followed illustrated their nuanced views and range of experiences (Kochel, 2014). Many residents, particularly minority residents, participated in the peaceful protest that followed Mr. Brown's death. Most residents disagreed with the police response to civil unrest, finding it over-aggressive. African American residents were more likely to strongly disagree with police response than non-Black residents. Residents also disagreed with the public vandalism, rioting, and looting that occurred. Residents reported negative experiences ranging from relatively minor inconvenience from disruption of shopping or travel habits, due to road or store closures, to suffering personal injury from violence or exposure to tear gas, to having their post-traumatic stress disorder triggered. Some had negative emotional responses, or lost trust in the police and their community. Yet others found a silver lining in these events, stating they had deepened their understanding of issues related to profiling, crime, and racial tensions or that they had become active in the community to counter injustice.

In summary, the literature on justice demonstrates that individuals experience justice in a social context, expecting to be treated with dignity and respect, and to experience procedural fairness from authorities. They infer not just their individual position in society, but the relative status of their social group. When these expectations are not met, it weakens part of the basic fabric of society, shaking faith in justice and the law.

Social movements

This research is not specifically focused on a particular social movement or organization, though the deaths of Freddie Gray and Tamir Rice have been powerful, mobilizing events for Black Lives Matter. Thus, the literature of *social movements*, groups or collectives organizing and networking to promote (or resist) social change (Stewart et al., 2012c) may provide some perspective on the emerging and evolving local networks of justice-related discourse in Baltimore and Cleveland in the aftermath of the deaths of Freddie Gray and Tamir Rice. Social movements are substantially composed of actors who are not formally part of the institutions they seek to change, but instead are everyday people. Because social movements lack formal power, they may rely on mechanisms such as persuasion and communication to advance their goals. Social movements also use collective action, such as protests, to voice their grievances and concerns and demand change (Snow et al., 2008).

Social movements are typically observed to move through stages, though they can become quiescent or revert to an earlier stage (Stewart et al., 2012d). The *genesis* stage, when a movement begins, marks a time when some pre-existing problem or issue begins to garner more attention, perhaps coupled to a triggering incident. If a movement does not stall out at this point, it may advance to the *social unrest* stage. At this stage, the movement may formulate an ideology, organize gatherings, and garner attention from those institutions that it challenges as well as from the media. In addition, in the *unrest stage* the movement strives to persuade the public and institutions of the need for change while attracting more adherents. It does this by re-framing how an issue or situation is perceived. Snow and colleagues describe three types of framing that aid social movements in making progress towards their goals and in generating collective action (Snow et al., 1988). *Diagnostic framing* identifies the "problem" ailing society, who is being victimized by it, and who or what is responsible for causing the problem. Thus, it defines what or whom the social movement must stand against. *Prognostic framing* proposes solutions and tactics for resolving the problem. These are actions that the social movement must perform or changes that authorities must carry out. Finally, *motivating framing* rallies and inspires members of the movement to action.

As members and adherents lose faith that appeals to institutions utilizing standard societal mechanisms (e.g., petitions) will work, the movement may transition to the *enthusiastic mobilization* stage (Stewart et al., 2012d). In this stage, mass marches, demonstrations, or acts of civil disobedience increase pressure on the establishment. Celebrities and public figures may become involved, further legitimizing the movement. While progress can occur during the enthusiastic mobilization stage, it may be less transformative than desired.

As this level of activity and energy may not be sustainable, the movement will transition to the *maintenance* stage, in which the movement must sustain itself, defend its gains and strive for more. In this stage, the movement becomes more like an organization than an un-institutionalized collective. The *termination stage* marks the end of the movement. This can occur because its objectives have been achieved and institutionalized, or have been overtaken by events. Alternately, the movement may fizzle out, or it transform into another form, such as a pressure group.

The language of a movement, the content it creates and discourse it generates, may support identification of members with the movement, a united “we” vs. an external opponent. Movements will develop their own slogans or pithy phrases as signals (Stewart et al., 2012a).

Institutions coping with social movements may choose among multiple strategies when determining how to respond. Early on, it may be easiest to avoid or ignore the movement (Stewart et al., 2012b). Later, an institution that is challenged may respond by impugning the movement or trying to make it frightening to the populace. Institutions can coerce members of social movements into compliance via restrictive policies, the police, and similar methods. They may adjust to the demands of the social movement, making new agreements (which may or may not be followed), or co-operating with members, inviting their participation on boards, committees, and so on.

Becoming personally involved in protest or collective action has been associated with several social and psychological factors (Klandermans, 1996). A sense of grievance, perceiving an injustice or feeling that one's interests or principles are under attack, is one factor that affects engagement. The perception of efficacy, i.e., that action can lead to a result, is another factor. Additionally, changing one's social identity to associate with a group or movement, and the emergence of collective identity are also associated with personal involvement. Emotions triggered while appraising a situation or event with respect to the group, such as anger, can motivate protest. Finally, social embeddedness, being positioned within a network of others, not only exposes one to more motivating or mobilizing messages, it poses a relational context in which cooperative action is encouraged.

Thus, we can see that social movements and collective action are far more than simple instrumental responses to obtain a desired outcome. Rather, they involve complex activities and changes in the social and psychological perception of members, such as changes in identity. Social movements and collective action can even lead to changes in the views, values, beliefs, and behaviors in the general population.

Social media

Social media is bringing pervasive changes to many facets of modern society. Social media's affordances, reach, and low barriers to entry have contributed to massive worldwide adoption (Kaplan & Haenlein, 2010). Social media, and in particular social network sites like Twitter, allow individuals to define and present an identity, visibly connect or affiliate themselves with others, interact, and share content. Social media enables individuals to maintain and strengthen friendship and kin relationships, and to form new relationships (Ellison & boyd, 2013). It can help break down barriers of geography, age, gender, ethnicity, political affiliation, socioeconomic status, nationality, and more. In terms of health, social media has been associated with promoting healthcare quality for patients, increasing awareness of health-related information, providing social support, and improving patient outcomes, as well as supporting public health surveillance (Dredze, 2012; Moorhead et al., 2013).

Social media is driving a reshaping of social organizing (Juris, 2012). This includes decentralization of organization and activity, and more expansive outreach efforts. Virality, and greater ease of sustaining action, also occurs through social media. Social media has become a premiere news and information sharing medium (Kwak et al., 2010), and its utility in crisis and disaster response has been well documented (Heverin & Zach, 2010; Vieweg et al., 2008; Yates & Paquette, 2011). It can be used for political and public opinion mining (O'Connor et al., 2010), and even for providing insight into personality (Golbeck et al., 2011) .

Using social media as an information source has implications for scholarship as well. The ability to gather and analyze vast amounts of data generated by social media will be transformative to social science, and is enabling the development of a new field of computational social science (Lazer et al., 2009; Watts, 2007). Yet most theories of human social behavior and interaction were developed at a time when terabytes of "digital trace" data had not been conceived of, posing potential challenges for usage.

In the United States, social media has been adopted by over 90% of online adults (Perrin & Anderson, 2019). Roughly a quarter of this population uses Twitter. Males and females are equally likely to use Twitter, though users skew slightly younger and better educated than the general population.

In summary, social media, which simultaneously allows individuals to present themselves and to connect and interact with others, both reflects our changing society and helps drive its evolution. It can promote collective action and organization as well as supply the digital trace data that is enabling new scholarship linking computational and social science methods.

2.4 Social media and justice, policing, and political movements

According to Gordon and colleagues (2013), social media is playing a role in shaping the nature of civic engagement in the digital age. The foundations of civic engagement can be summarized as being able to acquire and process information, to voice one's opinions and debate with others, and to take action in a context of social or political institutions. The affordances of social media support all three. They can amplify tendencies for social interaction, provide a personalized experience of information and interaction, and reconfigure our sense of place or location. Adoption of online technologies can facilitate offline collective action, simplifying organization and mobilization, and extending reach even transnationally, while simultaneously creating new modes of digital collective action (Van Laer & Van Aelst, 2009). Social media also affords its users simple mechanisms to publicly signal group or collective identity or affiliation through adoption of visual indicators, such as profile images (Vie, 2014).

Social media can serve as a resource that allows marginalized groups or individuals to develop and amplify their perspectives, thus democratizing public discourse and recasting the framing of galvanizing events provided by mainstream media. Particularly on Twitter, the use of hashtags to construct, maintain, and transmit alternative framings of contested events, such as the death of Freddie Gray, has been observed (Welles & Jackson, 2019). On Twitter, this activity can be viewed as creating "networked counterpublics" and spawning opportunities for influential community members to emerge and broker mutual understanding around divisive issues or events. Twitter's affordances, including the ability to include multiple hashtags per tweet that target or attract different politically or ideologically oriented audiences, increase the visibility of and promote exposure to content that might otherwise go unseen (Conover et al., 2011).

Focusing on the Twitter temporal signatures and evolution of Occupy Wall Street—the 2011 movement protesting social and economic inequality—Conover (2013) observed that those actors who were most vocal early in the movement appeared to become less engaged in discourse and drift away over time. This research explored social media participants in the Occupy movement at a global level, rather than in specific locations where #occupy protests took place, such as New York City. It used hashtags relating to the #occupy movement to identify relevant tweets. This work confirmed that social media activity rates paralleled events on the ground, rising dramatically during protests, and dipping afterwards. By examining the tweet history of a random subset of 25,000 users who had produced at least one #occupy tweet over the 15-month period from June 2011 through August 2012, Conover's research found the proportion of #occupy-related content fell over time even for the most committed, early members of the movement. Over 40% of users from this subset had an average 64% of their tweeting activity dedicated to #Occupy a during the peak activity for the movement. By the last 3 months of the time period, fewer than 5% of them produced any #Occupy content at all (Conover et al., 2013).

A movement more closely associated with the issues of justice, police brutality, and racial inequality, #BlackLivesMatter, has attracted the attention of social media researchers. Work by Olteanu et al. (2015) characterized participation in #BlackLivesMatter discourse in demographic terms, crowdsourcing annotations of race, age, and gender using the Crowdflower platform. Judgments were based on user profile information. It assessed whether users were only active during event-related peaks, which were associated with the killings of African Americans by police. In this work, a "peak window" was defined as a four-day interval including the day of the peak, the day before the peak, and two days after the peak. This research also determined whether participants were involved at a high, moderate, or low activity level, based on their tweets. It found that African Americans were overrepresented, given their proportion of the population, in engagement with #BlackLivesMatter, and were also the most active participants.

Using Twitter as a "sensor" to relate online activity to protests on the ground, De Choudhury et al. (2016) examined participation in #BlackLivesMatter from a temporal and geographic perspective, delving into how engagement and language relate to both historic rates of violence against African Americans and current protest activity. This research found that while the movement continued to attract more participants on Twitter, ongoing participation from established participants also occurred. Using data collected for a month-long period pertaining to each of several incidents (in Ferguson, MO, New York City, and Baltimore, MD) containing relevant hashtags, de Choudhury and colleagues found that the proportion of new participants peaked during important events with high volumes of activity. Using Linguistic Inquiry and Word Count (LIWC) to quantify tweet text, differences in identity-related LIWC variables were correlated with future protests on the ground. These include both a decrease in first person singular ("I") pronouns and an increase in first person plural ("we") pronouns the day before a protest. This was interpreted as evidence of an increase in social identification with the movement.

Social media activity has been linked to other forms of political action. Vissers and Stolle (2014), in a two-wave panel study of Facebook users, found Facebook political activity in Time 1 seemed to mobilize offline protest activity in Time 2. They also observed the reciprocal relationship, concluding that political expression on social media and offline protesting are interwoven phenomena. Further, they suggested future work considering network structure could help illuminate how social media activities and protests reinforce each other. Likewise, in studying Facebook use in the months leading up to the 2008 presidential election, Vitak et al. (2011) found a strong positive correlation between users' political activity on Facebook and their offline political engagement, and that those engaging in high levels of political posting on the site also had more peers engaging in similar behaviors (Vitak et al., 2011).

The relationship between social media and political protests has also been framed in the network terms of "strong" (friends and family) vs. "weak" (acquaintances) social ties, with Facebook networks seen as representing more strong ties, such as those between friends and family, while Twitter represents networks of primarily weak ties (Valenzuela et al., 2014). This study also adds a temporal consideration to its analysis. It anticipates that online social networks will be most crucial in the early stages of movements, enabling information acquisition and social coordination, and in later stages, when the protest movement is in decline. During the movement's peak, news outlets and other traditional media may effectively serve as primary information sources. In Valenzuela and colleagues' study, regular Facebook users and Twitter users both showed increased participation in protests over individuals who didn't use these social media platforms regularly.

Voting behavior, another type of political action, has also shown social transmission in social media via strong ties (Bond et al., 2012). In a randomized trial of tens of millions of Facebook users, users exposed to "I voted" messages coupled with profile pictures of their close friends on Facebook (the "social message" condition), showed a small but significant increase in real-world voting, as well as online information seeking about polling locations. Users in the control group, who were exposed to an informational message about voting with no social component, did not show an increase in real-world voting.

To summarize, social media is enabling collective action and is shaping social identity and identity presentation in new ways. Analysis of social media has demonstrated that online social media activity can translate into physical world actions in the political sphere. By revealing patterns of activity and content indicators, researchers are increasingly able to characterize activist movements and even predict protests.

2.5 Social networks, justice perceptions, and collective action

Social networks are sets of actors who are connected or tied to each other through a relationship, such as friendship, kinship, communication, or support. Social network analysis (SNA) is an approach that is fundamentally relationally driven, focusing on ties between actors rather than on the individual's attributes or properties (Borgatti, Mehra, Brass, & Labianca, 2009). SNA holds that network structure and actor position within a network are inherently valuable objects of inquiry. As a consequence, networks may serve as a conduit for opportunity and resources, providing paths through which these things flow. Alternately, networks may enmesh and restrict the actions of actors, who are embedded in a web of ties to others. Networks may transmit innovations, or disease. Observed properties and structure of human social networks are not random. Instead, they reflect underlying social processes and generate emergent social phenomena.

The study of social networks has contributed to our knowledge in areas as diverse as the diffusion of innovation (Valente, 1996), social capital (Wellman et al., 2001), the structure and performance of organizations (Borgatti & Foster, 2003), and health and disease (Christakis & Fowler, 2007). For an individual, expanding one's network through the formation of new weak ties to others in particular can provide novel information or similar benefits typically not obtainable through close ties with intimate or familiar partners (M. S. Granovetter, 1973).

The social network perspective moves beyond the individual actor to consider indirect paths of influence and global network structure as shaping behavior, feelings, and beliefs (Shapiro et al., 2008). Networks provide a structure through which attitudes and perceptions about justice may diffuse, be amplified, modified, or sputter out. Actors who have experienced an injustice may share their experiences and perceptions with those in their social networks, and reach out for social support as well. Those network members they have reached may, in turn, be influenced by these accounts of injustice and update their beliefs, even though they have not personally experienced any injustice. They may also transmit these perceptions and reports of injustice to additional people in the network. A densely connected, cohesive group can thus arrive at a collectively shared belief about justice based on injury to one or a few members.

Because people do not operate in a vacuum, but live, learn, and act within their networks, insight into network structure may inform collective action. Siegel (2009) examined how the structure of networks could affect aggregate political outcomes, and how network structure interacts with individual motivations to drive political participation. Through simulating basic network structures such as random networks, rings, hierarchical networks, stars, and small world networks, as well as varying the motivation of actors to participate, Siegel's research established that the overall structure of the network is an essential component of behavioral spread and participation rates. In his model, actors' motivations to participate were weighted by the participation of those in their local neighborhood.

Siegel (2009) also found complex dynamics emerged across different network structures. Identical distributions of motivation can produce different results. For example, a small world network structure will typically enable spread—and thus political participation—more effectively than a comparable hierarchical network. Even the presence of highly connected, motivated actors cannot always overcome the "roadblock" posed by a network's structure.

Social networks can play a role in individual decisions to participate in a social movement, with strong, pre-existing social bonds predictive of participation. Networks influence both the individual and the collective behavior of a movement (Diani, 2008). Network properties, such as centralization, appear to be associated with a higher propensity for activating collective action. Centralization is the tendency for network ties to be concentrated or "centralized" in a small number of actors.

While the literature on social network structure and justice-related collective action is not extensive, there may be insights to be drawn from studies of other challenging social dilemmas. Natural resource management is one such area. Environmental, social, and economic positions on management of a natural resource such as a fishery may be entirely at odds with each other (Conley & Moote, 2003).

Bodin and Crona (2009) examined the role of networks in natural resource management. Social networks can have powerful effects on social processes such as information sharing, consensus building, and power, all of which come into play when managing ecosystems. The authors found that network-level properties can make a difference in how successfully stakeholders navigate their challenges. For example, increased network centralization was correlated with collective action, as a centralized structure can simplify coordination and direction. But there was not a simple linear relationship between centralization and success. A network optimized for mobilizing and coordinating actors may be ideal in the early stages of natural resource management, when promoting awareness and engagement are critical, but suboptimal during mature stages of governance, when complex tasks involving integrating the perspectives of peripheral actors dominate.

As is true with much social science work, collection of appropriate, accurate, and sufficient network data is often challenging (Marsden, 1990). Some relationships may be observable in the public sphere, captured by social media or other information and communication technologies (ICTs), though naïve extrapolation from recorded data and metadata has limitations (Howison et al., 2011; Lazer et al., 2009).

To summarize, social network analysis (SNA) is a well-established field of scientific inquiry. Social networks are constructed of ties that connect individuals—bonds of kinship, friendship, collaboration, social support, and more. Networks enable the flow of ideas, perceptions, and beliefs. Social networks can play a powerful role in shaping individuals' actions and identity, as well as their attitudes about justice. Social networks can also support and sustain collective action. ICTs can be a powerful source of social network data at scale, if appropriately interpreted. For this research, network factors are essential to understanding ongoing levels of engagement with justice on social media after a salient event.

2.6 Dissertation Research Questions

This chapter has reviewed literature concerning issues of justice, collective action, and social movements, and factors that contribute to ongoing engagement in such efforts. For this dissertation, the main factors of interest are social identity, emotions triggered by the issues and representative events, perceived efficacy or ability to take an action and be heard, and social embeddedness, or relationships with others involved with the group or issue. Additional literature indicates that social media seems to be reshaping the nature of civic engagement in society, with digital engagement becoming a prominent feature of the landscape of social and political discourse and change.

While some social media research has associated changes in specific variables (e.g., "I" vs. "we" pronoun usage as an indicator of social identity, or seeing posts about network alters voting leading to changes in voting rates), there is a gap in understanding whether and to what degree factors involved in predicting physical world engagement are also associated with digital engagement. Building on this literature, this dissertation evaluates a set of factors that may predict Twitter users' ongoing engagement in justice discourse following a galvanizing event. Using two large datasets of tweets from Baltimore and Cleveland following the police killings of a young Black man and a 12-year-old Black boy, analyses focus on the following four research questions:

RQ1. Do Twitter users' expressions of identity in post and profile content predict ongoing digital engagement?

RQ2. Do Twitter users' expressions of emotion in post content predict ongoing digital engagement?

RQ3. Does generating more effortful posts on Twitter predict ongoing digital engagement?

RQ4. Does a Twitter user's social embeddedness predict ongoing digital engagement?

In the following chapter, I provide details on the two datasets collected for these analyses, as well as how the various independent variables and dependent variable (ongoing engagement in social justice discourse) were computed.

Chapter 3: Approach and Methods

3.1 Overview of Approach

This research explores and advances understanding of how a social media-enabled community concerned with issues of justice evolves and develops in the aftermath of highly salient local events, such as the death of Freddie Gray in Baltimore. This research develops a model to predict who, among those who newly engage with the topic of justice on social media during a salient event, will stay engaged over time. It employs the dimensions of *identity, emotion, social embeddedness*, and *effort*, adapted from prior literature, as the foundations of the model. Leveraging logistic regression, this approach produces a relatively interpretable model to help quantify the degree to which each of these dimensions contributes to prediction of ongoing engagement. After developing an approach to this case, the same approach is applied to a second event, the shooting of Tamir Rice by police in Cleveland, to demonstrate its generalizability. This dissertation extends prior work focused on social media posts associated with specific activist movements by examining posting relating to the fundamental concept of justice within a community.

Discussions of justice on social media are not surprising in the wake of events as wrenching as the death of Freddie Gray and the riots that followed. For some, these discussions are simply short-lived signals of interest, and will quickly fade, just as reporting on the Baltimore riots faded from the national news. Alternately, these discussions may serve as indicators of a more persistent change within certain members of the community.

The first dataset explored is a large corpus of tweets from Baltimore, Maryland around the time of the arrest and death of Freddie Gray while in police custody (April 2015), an event that shook the city. From that corpus, relevant tweets and users were identified using machine learning methods. For this phase of the work, I built multiple classifiers to identify justice-related tweets, then created an ensemble model from those classifiers. These classification results identified the set of relevant tweets from within the tweet corpus to be used in the next stages of this work. In contrast to methods that collect data based solely on relevant hashtags, this work focuses on a geographically bounded community (people identifying as being in Baltimore), rather than considering global discourse on the topic.

Once those relevant tweets were determined, this curated dataset was used to supply context for interpreting the findings, and to generate user-level aggregated data for the predictive model. For example, each justice tweet had a score from the police image model, which was null if the tweet had no image, and ranged from 0-1 for tweets containing an image. At the user level, this information was mapped to a variable that was the count of user tweets that scored > 0.5 for containing a police image. More than 60 variables were computed for each user to support prediction of continued engagement over time. The importance of the dimensions of *identity*, *emotion*, *effort*, and *social embeddedness* to predicting engagement is examined by considering the user variables in the digital trace data that can be reasonably mapped to each of these dimensions through logistic regression models. The steps of this process are captured in Figure 1 below.

To enable this research, I established a baseline for justice-related social media discourse participants in Baltimore. Data for this initial phase (T1) includes relevant social media up to the day of Freddie Gray's death. The second phase (T2) covers the intense period of protests, rioting, and violence that followed. The final time period (T3) covers post-riot activity. Comparable phases were established for the Tamir Rice incident.

3.2 Twitter data from Baltimore and Cleveland

Data was collected by a query to Twitter's Historical Powertrack[1], a service that will return all tweets in the Twitter database that match the query's criteria. A bounded spatiotemporal dataset was most appropriate given the research questions.

A query to retrieve data from the time period preceding the arrest of Freddie Gray through the weeks following his death from the Baltimore area was constructed. The query returned all tweets between April 4, 2015 and May 29, 2015. The tweets returned by the query were all geocoded to the city of Baltimore. They include geo-tagged tweets as well as tweets that were geo-located via a place check-in or via the location supplied by the user's profile. While geo-located data is not a perfect representation of physical location, it has been observed to be roughly 90% accurate in reflecting physical location (Burton, Tanner, Giraud-Carrier, West, and Barnes,

[1] More details on Historical Powertrack are available at http://support.gnip.com/apis/historical_api/overview.html.

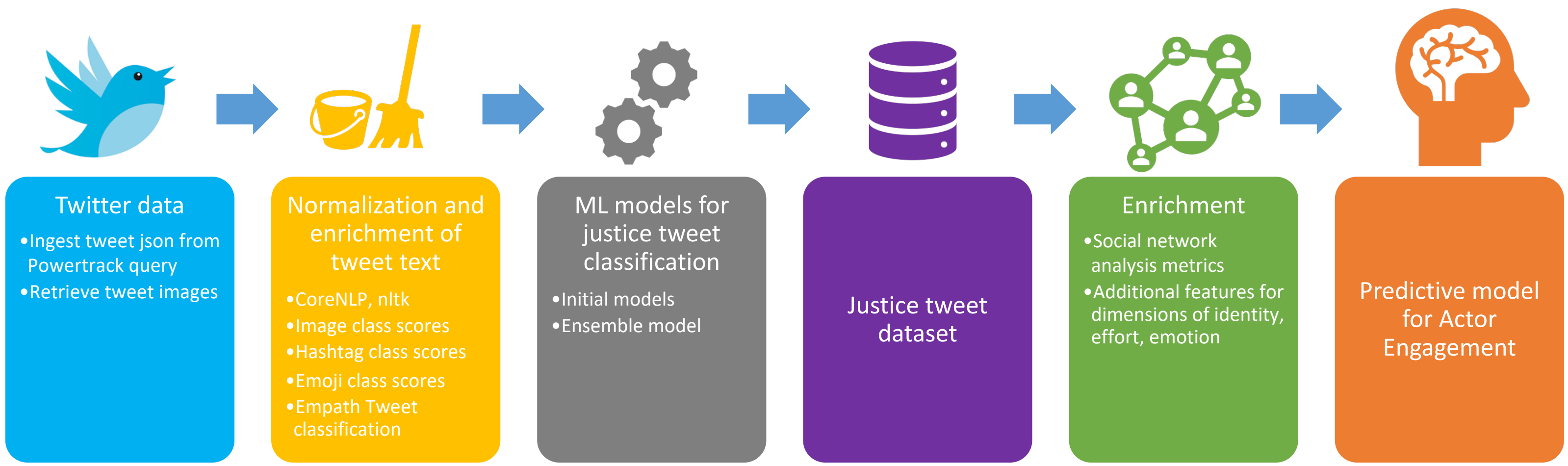


Figure 1: Overview of tweet processing and classification approach to support analysis. Tweets are ingested and their associated images are retrieved. The tweet text undergoes several processing steps to improve its usefulness for classification. Additional enrichment features are computed. All of these features are available to be used in the classification step. After classification, a dataset of relevant, enriched tweets is produced for further analysis and the predictive model of user engagement in justice discourse.

2012; Dor et al., 2015), and captures a far larger percentage of Twitter activity and users than geo-tagging.

To address the research questions, tweets were partitioned into one of three time periods. The time period from April 4, 2015, up to the death of Freddie Gray on April 19, 2015, is the baseline of justice related discourse (T1), capturing the community of social media discourse prior to the salient event. The next time period extended from April 20-May 1, 2015 (T2). It begins the day after Freddie Gray's death, as news began to break, brackets the Baltimore riots, and ends with the cessation of the riots. This time period includes actors who joined in discussing justice in the context of this high salience event. The final time period (T3) begins on May 1, 2015 and continues until May 29, 2015, when data collection ended. Users who were active during T3 maintained engagement with the topic of justice despite the end of the rioting and decrease in media coverage in the city.

The Baltimore tweet corpus consists of over 5.9 million tweets from 60,706 user accounts. Of these tweets, 30% (1.8 million) were retweets. Over 1.3 million tweets (22%) contained hashtags. Slightly fewer tweets shared images (17%, or nearly 1 million). The city generated over 100,000 tweets on a typical day for this time period, though activity more than doubled, to over 200,000 tweets per day, during the riots.

As with the Baltimore tweet corpus, data was collected to build the Cleveland corpus by a query to Twitter's Historical Powertrack. A query to retrieve data from the time period preceding the shooting of Tamir Rice through the weeks following his death from the Cleveland area was constructed. The query returned all tweets between November 15, 2014, and February 15, 2015. This time period begins the week before Tamir Rice's killing to help provide a baseline. It

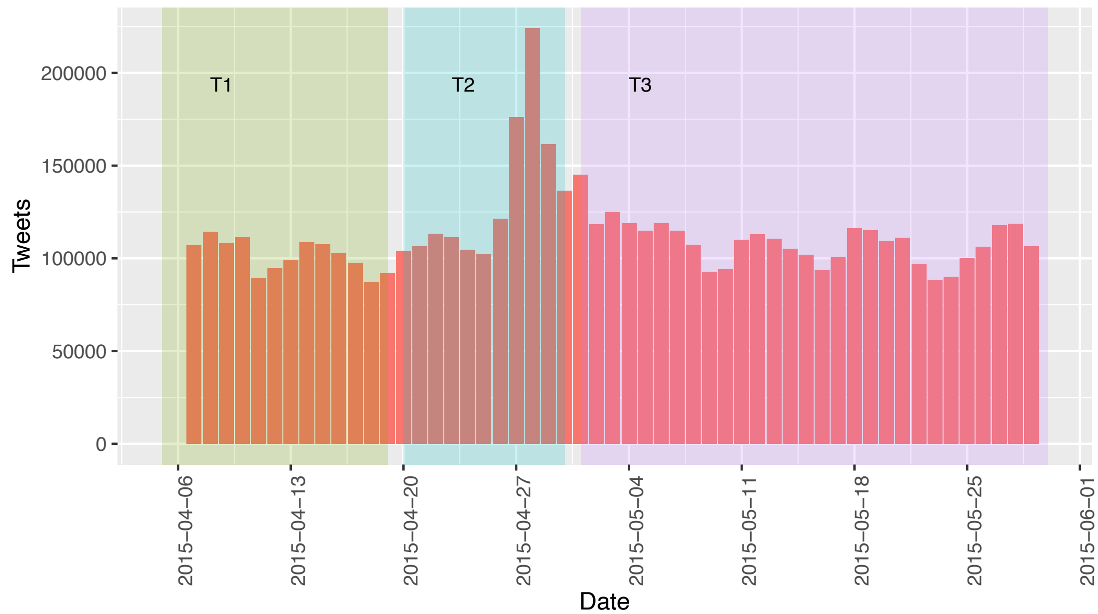


Figure 2: Tweets per day in Baltimore. From an average of about 100,000 tweets per day, activity spikes around the time of Freddy Gray's death and the violent protests in Baltimore.

extends past his funeral and protests over his death. As there were no rapid actions against the officers by the city, the timeframe extends longer than for Baltimore.

As with Baltimore, tweets were partitioned into three time periods. The time period up to the shooting of Tamir Rice (November 15 through November 22, 2014) is the baseline of justice related discourse (T1), capturing the community of social media discourse in Cleveland prior to the salient event. T2 extends from November 23 to November 28, 2014. It begins the morning that Tamir Rice's death is announced, and includes the peak of protest on November 25 (coinciding with the grand jury decision not to indict in the indict police officers in the separate shooting of Michael Brown in St. Louis), the release of surveillance video on November 26 from Tamir Rice's killing, and ends on November 28, after five consecutive days of public protest. This time period includes actors who joined in discussing justice in the context of this high-salience event. The final time period (T3) begins on November 29, 2014 and continues until February 15, 2015. Users who were active during this time period (T3) maintained engagement with the topic of justice despite the end of the protests and substantial decrease in media coverage of the events. Cleveland did not experience rioting or the deployment of the National Guard, nor swift action from the city to announce charges, but experienced a prolonged period without official closure. Tamir Rice’s death was not declared a homicide until mid-December 2014, and the investigation was transferred to an outside jurisdiction in mid-January 2015. At the end of January, the Rice family amended their lawsuit against the city to include negligence, Cleveland police's pattern of use of excessive force, and deliberate indifference to a serious medical need. The end of T3 for Cleveland was established as one month after the criminal investigation was transferred to the Cuyahoga County Sheriff’s Department. This approximates the one-month time span from criminal charges being filed in the Freddie Gray case to the end of T3 in Baltimore. Decisions on whether to indict or fire the officers dragged on for years. In April 2021, the family of Tamir Rice requested that the Justice Department reopen its investigation into the shooting.

The full Cleveland dataset contains over 8.4 million tweets from 60,851 unique users. The breakdown of tweets in this dataset were similar to the Baltimore dataset. Roughly 25% of tweets (2.2 million) were retweets. Images were present in nearly 15% of the corpus (1.2 million tweets), and 20% (1.7 million) contained hashtags. With a smaller population than Baltimore, the city generates close to 100,000 tweets on a typical day for this time period. During days of peak activity, this increased by more than 50%.

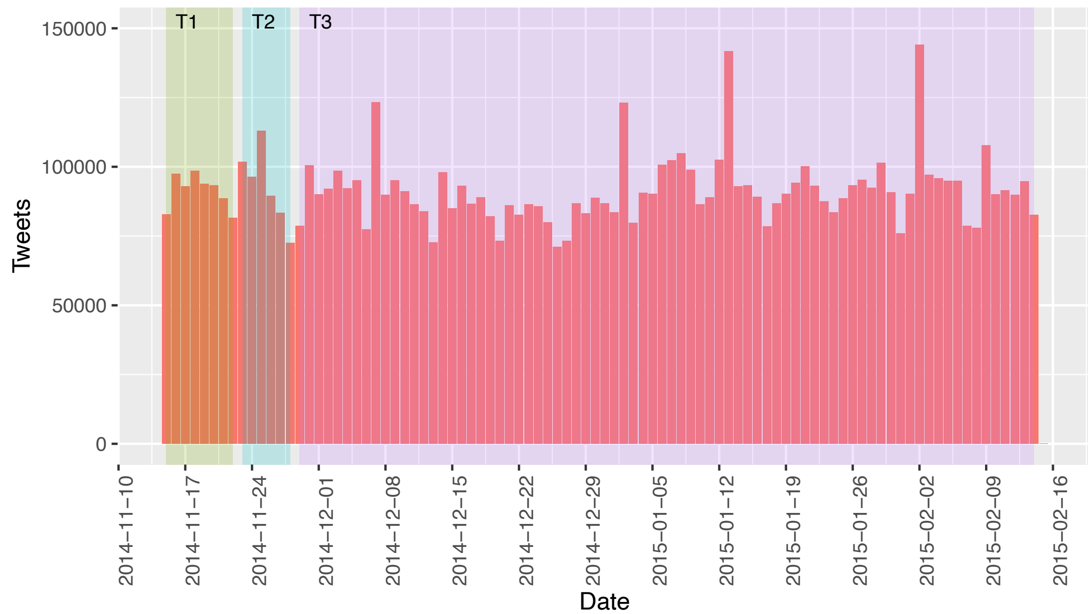


Figure 3: Tweets per day in Cleveland. While there is an increase in activity around the time of Tamir Rice's death, the spike is much less pronounced than was observed in Baltimore.

3.3 Processing tweet text and emoji

The text of a tweet is a primary means of conveying the tweet's information. Thus, incorporating text features in a classification task is a reasonable approach. In addition to its text, the body of a tweet may also contain non-text tokens, such as emoji, or entirely non-text elements, such as an embedded image. Since all of these elements can be used to convey meaning, I consider all of them as potential features in classification. This section describes processing steps applied to text and emoji. Steps applied to hashtags and images are covered in later sections.

3.3.1 Text processing

As an expression of natural language, the raw text of a tweet can be quite variable. In order to support machine learning, a text processing pipeline based on widely used natural language processing capabilities was used to enrich and normalize tweet text within the Cobalt system.[2] The text was processed using the Stanford CoreNLP natural language processing toolkit (Manning et al., 2014). The natural language toolkit was also used to process the text, producing stemmed and lemmatized features for the text, as well as a normalized text representation (Loper & Bird, 2002). As an example, here is the original text from a tweet in the Cleveland data:

Mother of #TamirRice expected to speak at any minute. Hear live coverage on WTAM 1100. Listen now http://t.co/I1bTKPj1sT

After processing, the normalized version the original text is:

mother tamirrice expected speak minute hear live coverage wtam listen

3.3.2 Empath

The tweet text from all tweets is also scored against the Empath categories, using the Cobalt system. Empath uses deep learning to construct dictionaries for a broad set of categories that cover a larger semantic space than LIWC (Fast et al., 2016). Beginning with seed terms for 200 categories, then using word embeddings to discover additional related terms for each category, Empath represents a data-driven approach to constructing lexicons. Empath demonstrated over 90% correlation with LIWC for LIWC categories when trained on a corpus of 1.8 billion words of fiction (Fast et al., 2016). Empath categories similar to LIWC include *positive_emotion, negative_emotion*, and *death*. Other Empath categories reflect concepts as diverse as *animals, crime, government, sport, technology,* and *vacation*. This large set of

[2] Tweet processing and enrichment was performed within the Cobalt system, a proprietary APL-developed system featuring highly scalable compute and storage, format-agnostic data ingest, automatic data normalization, custom text and image model training, a REST API with Swagger API specification, and a native Python library. Cobalt was designed to ingest and enrich over 500,000 tweets per day as well as the LexisNexis news feed and other information sources.

semantically diverse features is available for potential use in the justice tweet classification stage, or could later be leveraged for prediction of actor engagement. See Appendix C for a full list of Empath categories.

3.3.3 Emoji

There were roughly 1000 unique emoji in the datasets, occurring from over 600,000 times to just once in the data. Most emoji are used infrequently. As prior work has found that emoji can be emotionally laden (Novak, 2015), I anticipated that it may be useful to create features from this large set of items by grouping similar emoji into classes. Given the intensely polarizing and emotional nature of the events and phenomena we were examining—the deaths of a young man in Baltimore while in police custody and the shooting death of a 12-year-old boy in Cleveland, as well as the protests and riots that followed seeking justice—I anticipated that emoji that express emotion may be relevant features in the justice context. Similarly to the logic of sentiment analysis (Pang & Lee, 2008), I developed two classes of emoji relating to emotion. One class contained emoji expressing positive emotion, and another contains emoji expressing negative emotions. I also considered that emoji that depict people may be relevant, so I created a third class, as these might be indicators of reference to human or social phenomena.

While theories of emotion are numerous, and there is no fundamental agreement on precisely what the basic human emotions are, there are substantial overlaps between those theories, and little disagreement on positive or negative valence (Ortony & Turner, 1990) To operationalize emotion for this research, positive and negative affect categories are based on the work of Goleman (1995).The following table summarizes the positive and negative emotions.

Table 1: Positive and Negative emotions for Emoji annotation, drawn from Goleman (1995)

Type	Class	Examples
Positive	Enjoyment	Happiness, joy, relief, contentment, delight, amusement, pride, thrill, gratification, satisfaction, euphoria, ecstasy
	Love	Acceptance, friendliness, trust, kindness, affinity, devotion, adoration, infatuation
	Surprise	Shock, astonishment, amazement, and wonder
Negative	Fear	Anxiety, apprehension, nervousness, concern, consternation, misgiving, wariness, dread, fright, terror
	Anger	Fury, outrage, resentment, exasperation, indignation, animosity, annoyance, irritability, hostility, hatred
	Sadness	Grief, sorrow, cheerlessness, gloom, melancholy, self-pity, loneliness, dejection, despair
	Disgust	Contempt, distain, scorn, abhorrence, aversion, distaste, and revulsion
	Shame	Guilt, embarrassment, chagrin, remorse, humiliation, regret, mortification, and contrition

A codebook for emoji classification was developed. The coding guidance for emotion defined relevant emoji as those emoji that inherently express an emotion, versus one's personal affective response to the object depicted in the emoji. An initial subset of emoji was annotated by two annotators to validate the codebook. Few discrepancies were observed. For disagreements, the two annotators met and discussed to come to a final consensus, which could involve revising the codebook. For example, annotators initially disagreed as to whether heart emoji should be treated as Person (as body parts) if they were colored other than red, such as blue or purple. In seven additional cases, the emotive content or its valence was not obvious from the emoji itself. Emoji_sleeping_face was interpreted as positive, indicating a relaxed, peaceful state by one annotator, while interpreted as negative, expressing tedium and boredom by another. The emoji_bomb was interpreted as negative, expressing hostility and aggression by one annotator, while interpreted as positive (being "the bomb") by another. In these cases, a sample of tweets containing these emoji were retrieved from the corpus and reviewed to generate a final judgment.

All emoji that occurred at least 200 times in the complete Baltimore dataset were independently annotated by both annotators. For emotion, initial annotator agreement was high (see Table 2).

Table 2: Inter-annotator agreement for emotion classification of emoji. Two annotators assessed all emoji occurring at least 200 times in the Baltimore dataset for emotion, judging the positive or negative emotional valence of the emoji.

	Percent Agree	**Scott's Pi**	**Cohen's Kappa**	**Krippendorff's Alpha (nominal)**	**N Agree**	**N Disagree**	**N Cases**	**N Decisions**
Emotion	94.20%	0.849	0.849	0.85	440	27	467	934

An identical process was followed to determine person emoji. Person emoji are defined in the coding guidance as emoji that depict adults, children, professions with people, faces, or body parts (hearts, ear, etc.). This class does not include accoutrements of persons, such eyeglasses or clothing or jewelry. Annotator agreement is again high (see Table 3). There are 136 person emoji. A list of person emoji and positive and negative emotion emoji is available in Appendix A. All tweets in both datasets were enriched with a count of how many positive emotion, negative emotion, and person emoji they contained.

Table 3: Inter-annotator agreement for emoji classification for person class. Two annotators assessed all emoji occurring at least 200 times in the Baltimore dataset for representing a person.

	Percent Agreement	**Scott's Pi**	**Cohen's Kappa**	**Krippendorff's Alpha (nominal)**	**N Agree**	**N Disagree**	**N Cases**	**N Decisions**
Person	97.40%	0.936	0.936	0.936	455	12	467	934

3.4 Image classes and classification

Recent advances in Deep Learning have greatly simplified and advanced the accuracy of image classification for objects (Krizhevsky et al., 2012) and for types of scenes (Zhou, 2012; Zhou et al., 2014). I used an implementation of Deep Learning technology (Gifford et al., 2015; Rodriguez et al., 2014) in the Cobalt system to build classifiers for images contained within the tweets in the dataset. This implementation uses a pre-trained convolutional neural network (CNN) trained on Imagenet (1000 classes, approximately 1.3 million images) as a universal feature extractor, followed by a Random Forest that leverages the features for classification. This image classification approach was designed to require only dozens to hundreds of positive images to train a classifier on a new class, and supports rapid creation of image classifiers by non-experts in image classification.[3] These classifiers recognize relevant objects or types of activity or scenes, such as police, peaceful protests, or riots. Potential positive images for each class were downloaded from an image search to Google Images or Bing. Negative images, images that do not include the relevant category or type of activity or scene, were drawn from labeled images from Imagenet (www.image-net.org). All images used for training and testing classifiers were reviewed to determine that they were positive or negative for their class. All images used in classification were photos, not cartoons or other artistic renderings. These classifiers were applied to all the available images from the dataset, and their scores were

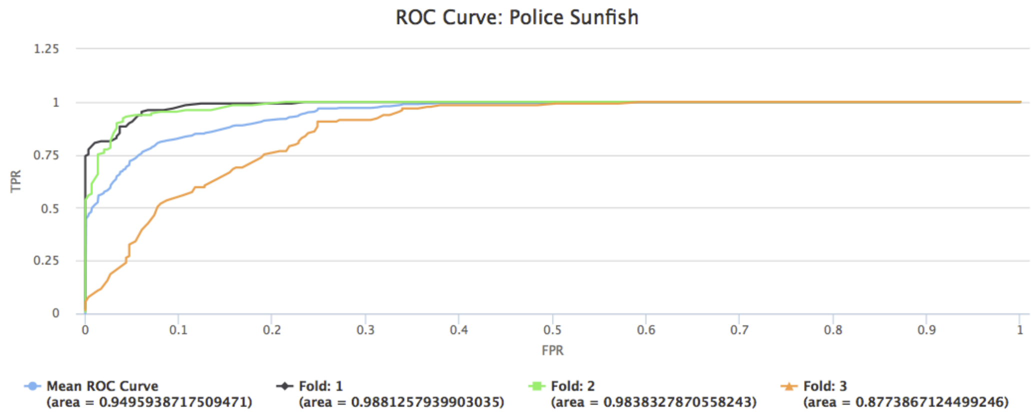
[3] This image classification system has been used to support several pilots and experiments exploring human mental models of image classification systems. For example, in one experiment, participants reviewed classifier outputs from a binary image classifier (bee/ non-bee) designed to have two flaws. The first was a false negative error where black and white photos of bees tended to classified as non-bees. This was instantiated in the training set by including very few black and white bee images and a number of black and white non-bees. The second was a false positive error where other insects appearing in conjunction with colorful flowers tended to be classified as bees. Participants articulated patterns of errors, and designed retraining sets by selecting additional images (e.g., black and white bee images) to compensate. After using these retraining sets to train new bee/non-bee classifiers, think-aloud participants reviewed the new results and classifier performance. The original training/test set for the bee/non-bee classifier consisted of just 550 images, evenly split between the positive and negative class.

included as tweet features for tweet classification. See Table 4 for the final list of image classes and descriptions.

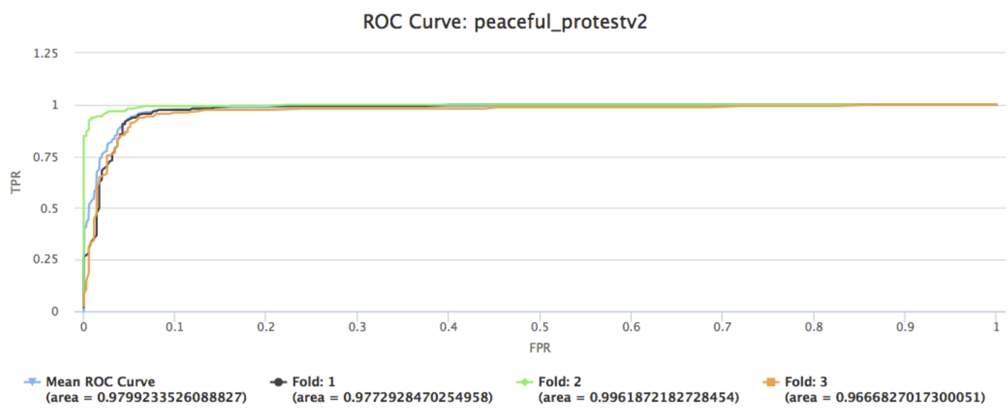
Table 4: Image classes and class descriptions. This guidance was used to assess images for use in creating Deep Learning image classifiers

Class	Description for coding
People	Multiple humans, any mix of adults and children.
Crowds	A large number of people gathered together, typically in a disorganized or unruly way.
Peaceful protest or march	An event at which people gather together to show disapproval about something; a mass demonstration in people attempt to make their opinions known publicly, in an effort to influence public opinion or policy. Can include picketing, marching, sit-ins, die-ins, etc. It does not include scenes of violence and destruction (not peaceful).
Parade	A procession of people, usually organized along a street, often in costume, and often accompanied by marching bands, floats or large balloons. Parades are held for a wide range of reasons, but are usually celebrations of some kind.
Police	Those empowered by the state to enforce the law, organized into forces. Uniformed police patrol, and respond to emergencies. Images will contain police officers, and may contain police vehicles or other police equipment. The police may be on patrol, performing their duties, or simply present and recognizable as police.
Rioting	A violent public disturbance against authority, property or people. Riots typically involve vandalism and the destruction of property, public or private. Images of rioting may include property damage and looting by rioters, protesters in violent conflict with police or officers in riot gear, fighting, arson or burning property (not a campfire, fireplace, or natural fire), crowds being fired upon or dispersed with teargas, etc.

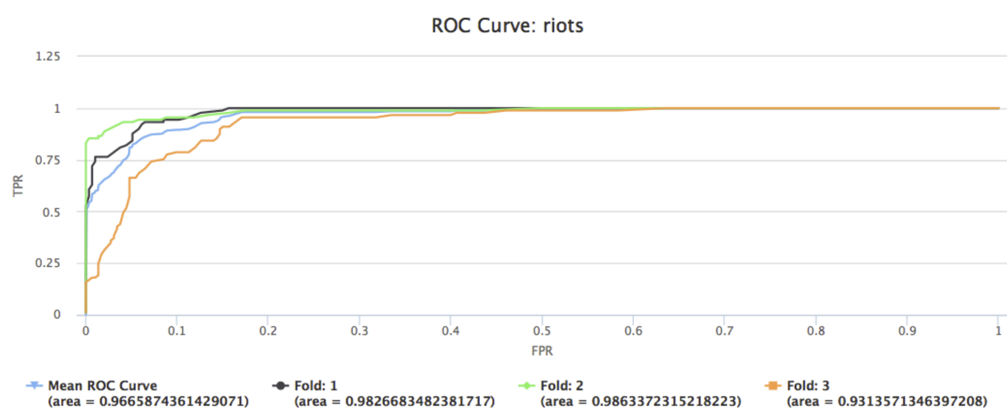
Image classifiers were created for these image classes, using 3-fold cross-validation. All of these classifiers performed well on test data, with mean area under the curve (AUC) above 0.9. For each of these classifiers, a receiver operating (ROC) curve and the distribution of positive and negative images is provided (see Figure 4). The scores (0.00-1.00) produced by each of these image classifiers were incorporated as features for each tweet that contained an image.



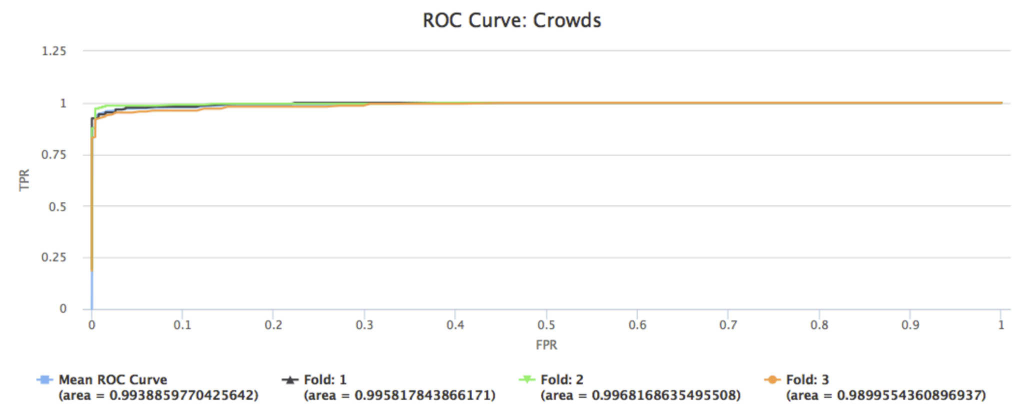
Number of Folds	Mean AUC	Positive Images Count	Negative Images Count
3	0.9495938717509471	387	895



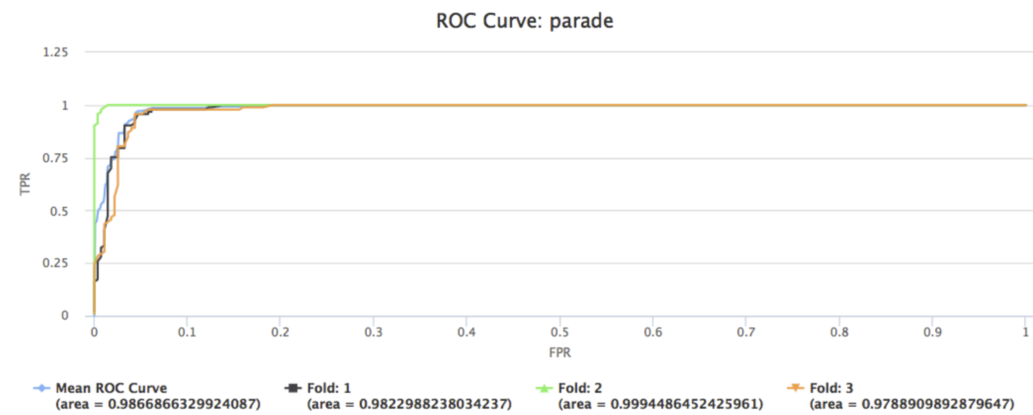
Number of Folds	Mean AUC	Positive Images Count	Negative Images Count
3	0.9799233526088827	478	1060



Number of Folds	Mean AUC	Positive Images Count	Negative Images Count
3	0.9665874361429071	267	881



Number of Folds	Mean AUC	Positive Images Count	Negative Images Count
3	0.9938859770425642	634	805



Number of Folds	Mean AUC	Positive Images Count	Negative Images Count
3	0.9866866329924087	278	829

Figure 4: Image classifier performance for all image classes. A Receiver Operating Curve (ROC) is shown.

3.5 Hashtags and hashtag categories

On Twitter, hashtags can be thought of as user-generated labels, tagging a tweet with a topic, or affective or evaluative response to the tweet's content. Hashtags enable discovery of and participation in topical conversations, rather than through following individuals. New hashtags are often coined in conjunction with events or issues (Bruns & Burgess, 2011). Further, hashtags concerning politically controversial topics have been found to be persistent, with implications for adoption (Romero et al., 2011).

Several hashtags were independently created, shared, and adopted in Baltimore relating to the events of April 2015 and the broader issue of justice. Community members may have chosen to emphasize different aspects of the events in their communication, or to signal different perspectives or affective responses to them through their choice of hashtags. For this reason, I created several hashtag categories to explore as potentially useful features for the Baltimore dataset. One class captures references to Freddie Gray (e.g., #FreddieGray), a second includes hashtags related to rioting (e.g., #BaltimoreRiots). A third category contains hashtags relating to justice itself (e.g., #JusticeforFreddieGray). Another category contains a small number of hashtags relating to the events in Baltimore more generally (#prayforBaltimore). These categories are not mutually exclusive. For example, *#justice4freddiegray* is a Freddie Gray hashtag and a Justice hashtag. For Cleveland, the second case analyzed in this study, similar hashtag categories for Tamir Rice and justice were created. As there was no rioting in Cleveland, there is no equivalent hashtag category.

3.5.1 Baltimore hashtags and hashtag categories

There were 286,320 unique hashtags in the Baltimore tweet dataset. Only a very small number relate to the issue of justice, or reference Freddie Gray or the riots that sprang up in the aftermath of his death. Yet several of the most frequently used hashtags in the corpus do relate to these events. See Table 5 for a list of the top hashtags used in the Baltimore dataset.

Table 5: Frequently occurring hashtags in the Baltimore tweet corpus. Hashtags pertaining to Freddie Gray, the rioting that took place after his death, or justice are **bolded**.

Hashtag	Count
Baltimore	92528
NowPlaying	55335

vape	41959
vapejuice	41919
vaporizer	41919
FreddieGray	**41034**
Orioles	28856
baltimore	22225
Job	19350
BaltimoreRiots	**17360**
BaltimoreUprising	**16076**
orioles	14468
nowplaying	14424
jobs	13856
listenlive	12825
Jobs	10560
etsymntt	9874
baseball	9586
MLB	8579
BlackLivesMatter	**5315**
OneBaltimore	5126
Birdland	4800
handmade	4715
job	4516
bigdata	4018

After reviewing the 1000 most frequent hashtags in the data, the hashtags in the table below were assigned to each class. Annotator agreement again was very high, with only a handful of disagreements. Many disagreements were borderline cases that one annotator was not particularly confident in. There was only one disagreement for the Freddie Gray hashtags, and two for Riot hashtags. One of the two annotators considered labeling #marilynmosby a Freddie Gray hashtag, because Ms. Mosby—the State's Attorney for Baltimore—was prosecuting the officers who were on trial for involvement in his death. In other cases, an annotator was unfamiliar with a hashtag for a proper name and did not recognize it was the name of an individual considered to have been a victim of police brutality (#TyroneWest). At least one of these relevant hashtags appear in about 100,000 unique tweets. There were roughly 50,000 tweets containing a Freddie Gray hashtag, 22,000 for riot, and 40,000 for justice. This hashtag count, particularly the justice count, could be considered a rough approximation for the "floor" of relevant justice tweets in the corpus. The frequency with which each hashtag, regardless of capitalization, was observed in the corpus is shown below (Table 6).

Table 6: Top relevant hashtags for Freddie Gray, Riot, and Justice. Of the thousand most frequently occurring hashtags in the data, a few dozen were relevant to justice, Freddie Gray, or the riots. Justice-related hashtags extended beyond calls for justice for Freddie Gray, evoking other victims of police violence, and larger social movements and actions.

Hashtags	Count	Freddie Gray	Riot	Justice
FreddieGray	43427	y		
BaltimoreRiots	18299		y	
BaltimoreUprising	16841			y
BlackLivesMatter	6183			y
JusticeForFreddieGray	2066	y		y
WalterScott	1709			y
BmoreUnited	1616			y
BaltimoreCurfew	1608		y	
BalitmoreUprising	1157			y
AllLivesMatter	917			y
Ferguson	904			y
dcRising	855			y
March2Justice	836			y
JusticeForFreddie	826	y		y
SayHerName	816			y
BalitmoreRiots	795		y	
RIPFreddieGray	762	y		
EndTheCurfew	761		y	
Justice4Freddie	611	y		y
FreddyGray	605	y		
baltimoreprotests	595			y
BaltimoreProtest	581			y
BreakTheCurfew	580		y	
BaltimorePolice	413			y
PeacefulProtest	413			y
police	390			y

BaltimoreJusticeFund	388		y
FreddieGrey	355	y	
TyroneWest	265		y
NoJusticeNoPeace	264		y
Justice4FreddieGray	255	y	y
policebrutality	171		y

It is notable that multiple hashtag variants appear for a number of identical concepts. For example, JusticeForFreddieGray, JusticeForFreddie, Justice4Freddie, and Justice4FreddieGray all convey an identical meaning, though they differ in form. This highlights one of the challenges of relying on hashtags to identify a corpus. Unless the specific hashtag token used is the focus of research, relevant content may be neglected.

In addition to this set of relevant hashtags annotated by two raters, to help address the issue of variants of relevant hashtags (often spelling variations or typographical errors), additional searches were conducted of the hashtags in the data, for relevant strings such as "fred", "gray", "police", "riot", and "justice." This detected several dozen lower frequency relevant hashtag variants (e.g., "Baltmoreriots"), which occurred a handful of times to a few dozen times. Each tweet was then scored for how many riot, Freddie Gray, or justice hashtags it contained.

3.5.2 Cleveland hashtags and hashtag categories

The same approach was applied to the Cleveland data, and additional sets of hashtags for justice and Tamir Rice were created for scoring tweets. There were 365,935 unique hashtags in the Cleveland tweet dataset. As was observed in Baltimore, only a few of the most common hashtags were related to the issue of justice or Tamir Rice.

Table 7: Frequently occurring hashtags in the Cleveland tweet corpus. Several pertain to justice or Tamir Rice. Most are not related to the topic.

Hashtags	Count
Cleveland	98517
Job	57262
Browns	44394
nowplaying	34657
CLE	33971
Cavs	33750
listenlive	28353
NBA	17303
NFL	14731
SoundCloud	13389
GoBucks	12479
crochet	12349
thisiscle	11754
Ferguson	**11620**
TNTweeters	11397
Buckeyes	8717
np	8566
RETWEET	8449
BTNH	8363
NBABallot	8251
Jobs	8099
GoBrowns	7597
UPDATE	7538
music	6799
FOLLOW	6690
MGWV	6643
TamirRice	**6487**
marketing	6440
tbt	6351
MLB	6329
NerveDJsMixtapes	6256
NerveDJs	6162
Ohio	6044
TeamFollowBack	5950
traffic	5874

Christmas	5871
gameinsight	5540
OHwx	5393
RT	5142
DeanAmbrose	5110
wkyc	5059
BlackLivesMatter	**4970**

Table 8 lists the hashtags identified as relevant to the justice or Tamir Rice categories in the Cleveland dataset. While there are a number of patterns in common with Baltimore, including hashtags that compound justice with the victim's name, references to other victims of police violence, and hashtags evoking social movements and slogans that are both supportive of substantial changes to the criminal justice system (#blacklivesmatter) and unsupportive of those changes (#alllivesmatter), it is notable that relevant hashtags are less common in this data than in the Baltimore data.

Table 8: Top relevant hashtags for Tamir Rice and Justice. Of the thousand most frequently occurring hashtags in the data, a few dozen were relevant to the topic. Justice-related hashtags extended beyond calls for justice for Tamir Rice, evoking other victims of police violence, and larger social movements and actions. Several refer to counter-narratives or movements supporting police, and downplaying systemic injustice and violence in the criminal justice system. Fewer relevant hashtags are identified in this dataset than were identified for Baltimore.

Hashtag	**Count**	**Tamir Rice**	**Justice**
ferguson	10452		y
tamirrice	6590	y	
blacklivesmatter	4979		y
ericgarner	2649		y
icantbreathe	1871		y
fergusondecision	1480		y
mikebrown	1097		y
riptamirrice	1009	y	y
chapelhillshooting	747		y
ferguson2cle	711		y

crimingwhilewhite	687		y
tanishaanderson	664		y
seaofblue	552		y
alllivesmatter	495		y
police	470		y
michaelbrown	418		y
darrenwilson	310		y

These hashtag features were then available for use in classification of tweets for justice, or for aggregation at the actor level and inclusion in predictive models of user engagement. I discuss this process in the next section. A list of the hashtags identified for each class is contained in Appendix E.

3.6 Tweet classification for justice

3.6.1 Developing a ground truth dataset for tweet classification

An essential component of this research is the ability to identify tweets that are likely to be topically relevant within corpora of roughly 6-8 million tweets. As this number of tweets is far too vast to manually classify, this research relied supervised machine learning. This involves annotating a sample of tweets to develop classifiers that can then be run against the entire corpus. Those tweets that score as positive for the relevant class then provide the data that will form the basis for the rest of the research. Classifiers for tweets pertaining to justice were developed and applied to both tweet corpora.

Once a set of ground truth data for training and testing had been created, the existing machine learning and text processing libraries of scikit-learn (Pedregosa et al., 2011) were used to build classifiers for justice related tweets. These libraries implement commonly used machine learning algorithms, such as support vector machines, decision trees, random forests, generalized linear models, and naïve Bayes. This was performed within the Cobalt system.

In order to establish a standard for this supervised learning task, a codebook had to be developed. The codebook helped ensure a clear, consistent standard for annotation. After the codebook was drafted, a sample of tweets was annotated by two annotators in order to detect inadequacies in the codebook that could be corrected. After this initial step of codebook revision, another set of tweets was annotated, to determine interrater reliability and validate the codebook's utility. The table below summarizes the coding guidance, which is available in Appendix B. User names mentioned in example tweets have been anonymized for non-public figures.

Table 9: Justice class annotation guidance for supervised classification for tweets. Example tweets that meet or do not meet the criteria for justice are provided.

Description	**Type**	**Examples**
Relating to any aspect of the U.S criminal justice system. It includes actions of actors in the system, events, policies and procedures, and responses by individuals or groups to any of the above.	Positive	#NoJusticeNoPeaceNoRacistPolice #BalitmoreUprising #BlackLivesMatter #BaltimoreProtests RT @ABC: Admire @SheriffClarke : #FreddieGray Charges 'Duke Lacrosse Case All Over Again' ://ow.ly/MqFDA #BaltimoreRiots http://t… RT @DEF: Y'all, Baltimore PD is putting out a narrative that #FreddieGray broke his on spine. https://t.co/JH1LiJN383
	Negative	Kobe had a longer career with more help. A.i. Averaged over 25 and 10 for 10 years straight literally spent the last hour arguing about feminism instead of writing my paper fantastic w/e Almost half of all Black children under the age of six are living in poverty. #PoorinAmerica

To generate a rough approximation of how many relevant tweets there were for each class, simple queries of the database on relevant terms for that class, based on terms from the hashtag classes were executed. This would help guide an appropriate sampling strategy to develop the ground truth set of annotated tweets. Queries on relevant hashtags for Baltimore returned fewer than 100,000 unique tweets, with approximately 49,000, 40,000 and 22,000 tweets containing hashtags relating to Freddie Gray, justice, and riots, respectively. A number of tweets contained hashtags from more than one class.

These results suggested that relevant tweets comprise only a few percent of the millions of total tweets in the corpus. Given the significant imbalance of the relevant tweets to the number of irrelevant tweets, randomly querying the dataset for tweets to annotate to train and test classifiers would return insufficient positive tweets. Instead, the terms based on relevant hashtags and their meanings were used as a basis to construct a query. These sets of terms were treated as a proxy to ensure a sufficient number of justice-related tweets are likely to be included in the dataset annotated for ground truth to construct the classifiers.

Randomly oversampling tweets with these features made it possible to address class imbalance. Four queries were constructed (one random, the other three based on the relevant sets of terms) and limited to 50,000 tweets. These terms (strings) included fred, gray, grey, polic, officer, justic, cop, arrest, riot, loot, fire, and curfew. Each of these results files was shuffled to randomize the order of the tweets, after which a sample of tweets was selected from each file and concatenated into a single file to constitute a set of source tweets for annotation.

In order to ensure sufficient information to evaluate the tweet, the URL for the original tweet was available to the annotators. Annotators could view the tweet, including any images, in its original context. This was particularly useful in cases where the text alone was ambiguous, as the example below illustrates. In the absence of the image, the text alone could only have been interpreted as unrelated to justice, but seeing the image depicting protesters invoking Freddie Gray results in a different annotation.

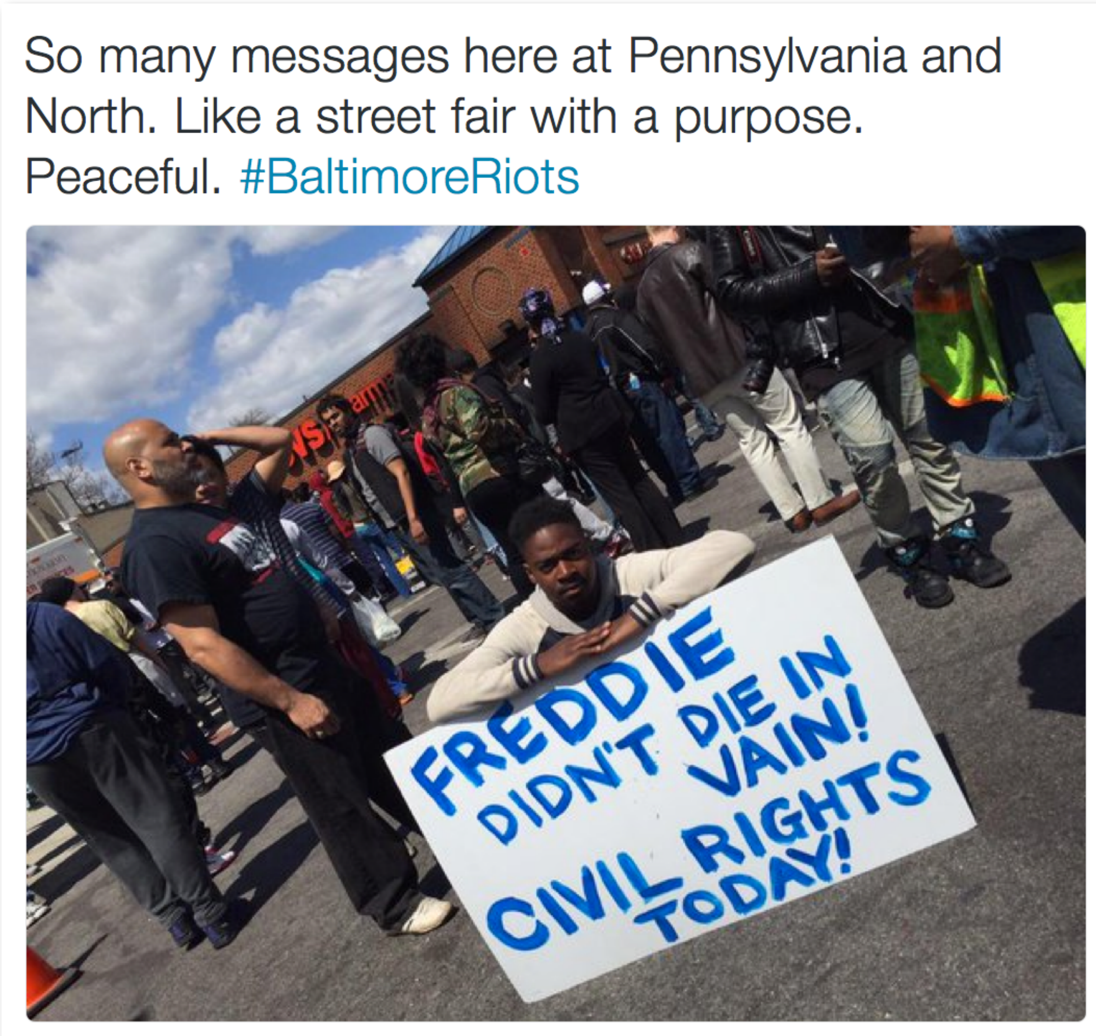


Figure 5: Example tweet containing an image that conveys information for justice classification

Initially, the two annotators marked 50 tweets, to familiarize themselves with the task and practice applying the codebook standards. This was followed by annotating 233 random tweets drawn from the concatenated file. Initial inter-annotator agreement was fairly high for the justice class, but could be improved. After review and discussion, the codebook was revised slightly to deal with this. This included clarifying that only tweets referring to criminal justice in the United States would be considered relevant. Reports of injustice in Tibet or other regions of the world would not be annotated as related to justice for the purpose of this research. The following table provides more detail on annotator agreement. After this step, additional tweets were annotated, to ensure sufficient data for building classifiers.

Table 10: Annotator agreement for justice annotations

	Percent Agree	**Scott's Pi**	**Cohen's Kappa**	**Krippendorff's Alpha (nominal)**	**N Agree**	**N Disagree**	**N Cases**	**N Decisions**
Justice	88.80%	0.725	0.726	0.726	207	26	233	466

Development of a ground truth dataset for Cleveland followed a similar path, though the refined codebook definitions did not need additional modifications. For both datasets, nearly two thousand tweets were annotated by one annotator, then reviewed by the other, with additional tweets being annotated by the author for ground truth. A total of 2462 tweets were annotated for Baltimore, and 2217 for Cleveland. In both cases, about 70% of tweets were in the negative class, and 30% in the positive class.

3.6.2 Tweet classifiers for Baltimore and Cleveland

A series of classifiers were built using tweets that were annotated for ground truth as described previously. Numerous combinations of features were considered, including the normalized tweet text (txt), images classification scores (img), emoji (emoj), relevant hashtags (tag), pronouns (pro), emotion (emo), and Empath features (death and crime). Models were computed in the Cobalt system, using scikit-learn gridsearch to tune parameters. Confusion matrices and additional measures of model performance are supplied in the following tables.[4]

Table 11: Confusion matrices for specific machine learning models on Baltimore justice data (binary classification of justice/not justice, n=2462, 80/20 split).

Model	**Confusion Matrix**		
	Actual	**Predicted**	
		0	1
JB2_balt_ent_txt_img_emo_tag_pro (Random Forest)	0	318	31
	1	33	111
JB@_Balt_txt_emoj_imtag_deathcrim (BaggingClassifier)	0	325	24
	1	28	116
JB2_txt_tag_img (Random Forest)	0	332	17
	1	47	97
JB2_txt (Random Forest)	0	339	10
	1	47	97

Table 12: Metrics of model performance for specific machine learning models on Baltimore justice data.

Model	**JB2_balt_ent_txt_img_ emo_tag_pro**	**JB@_Balt_txt_emoj_ imtag_deathcrim**	**JB2_txt_tag _img**	**JB2_txt**
Measure	Value	Value	Value	Value
Sensitivity	0.7817	0.8286	0.8509	0.9065
Specificity	0.906	0.9207	0.876	0.8782
Precision	0.7708	0.8056	0.6736	0.6736
Negative Predictive Value	0.9112	0.9312	0.9513	0.9713
False Positive Rate	0.094	0.0793	0.124	0.1218
False Discovery Rate	0.2292	0.1944	0.3264	0.3264
False Negative Rate	0.2183	0.1714	0.1491	0.0935
Accuracy	0.8702	0.8945	0.8702	0.8844
F1 Score	0.7817	0.8169	0.7519	0.7729

[4] Sensitivity (TPR = TP / (TP + FN)), Specificity (SPC = TN / (FP + TN)), Precision (PPV = TP / (TP + FP)), Negative Predictive Value (NPV = TN / (TN + FN)), False Positive Rate (FPR = FP / (FP + TN)), False Discovery Rate (FDR = FP / (FP + TP)), False Negative Rate (FNR = FN / (FN + TP)), Accuracy (ACC = (TP + TN) / (P + N)), F1 Score (F1 = 2TP / (2TP + FP + FN))

While performance of these classifiers on test data—an essential step for later analyses—showed utility, the value of very strong performance on the positive justice class to support the predictive user engagement model, as well as some variation in the number of total positive tweets from each individual classifier suggested an additional step would be helpful.

Prior work has demonstrated that combining the results of several classifiers in an ensemble or stacked generalization model can improve performance (Thorne et al., 2017). Out of the roughly 6 million total Baltimore tweets, 184,648 tweets were classified as positive by at least one model. In most cases, the majority of these four models agree on positive tweets about justice. The amount of agreement between each specific classifier and the other three is visualized in Figure 3 below.

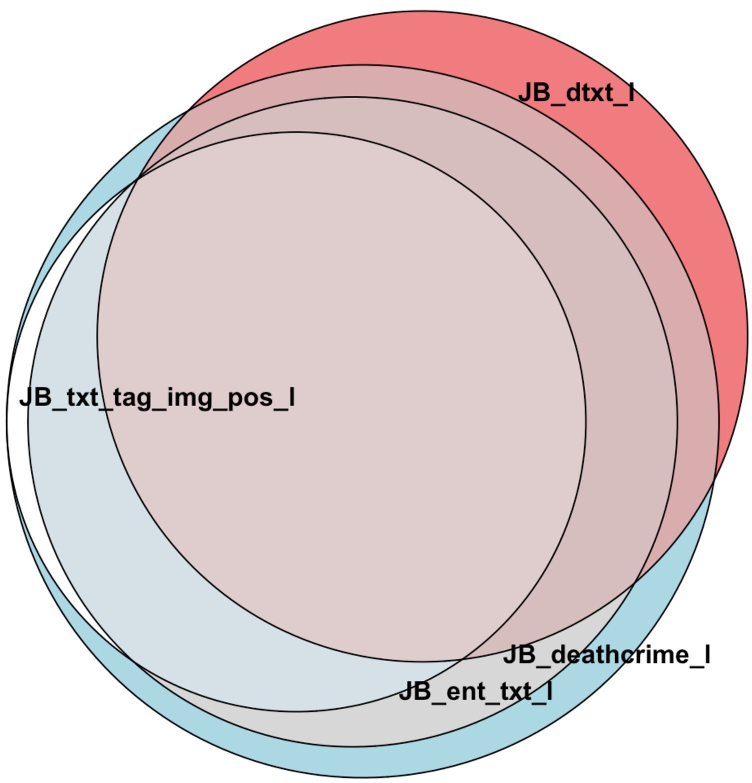


Figure 6: Overlap in agreement across Baltimore justice classifiers. Most tweets are classified positive by multiple classifiers.

Setting a threshold of two or more models in agreement produces a set of 158,738 positive justice tweets for Baltimore. These tweets were created by 15,881 unique users who tweeted about justice at any point during the timeframe. This method at this threshold improves performance in justice tweet classification.

Table 13: Performance of ensemble model for Baltimore justice classification data (binary classification of justice/not justice, n=2462).

Model	**Baltimore**
Measure	Value
Sensitivity	0.932
Specificity	0.985
Precision	0.962
Negative Predictive Value	0.973
False Positive Rate	0.015
False Discovery Rate	0.038
False Negative Rate	0.068
Accuracy	0.970
F1 Score	0.947

Following the approach used for the Baltimore dataset, multiple combinations of features were considered for Cleveland classifiers, including normalized tweet text (txt), images classification scores (img), emoji (emoj), relevant hashtags (tag), pronouns (pro), and Empath variables. Models were computed in the Cobalt system, using scikit-learn gridsearch to tune parameters. Confusion matrices and additional measures of model performance are supplied in the following tables.

Table 14: Confusion matrices for machine learning models for Cleveland justice data (binary classification of justice/not justice, n=2271, 80/20 split).

Model	**Confusion Matrix**		
		Predicted	
	Actual	0	1
	0	281	9
JC2_txt_tag_img (SGD)	1	8	146

	0	273	17
JC2_txt_img_pro_tag_emoji (Random Forest)	1	20	134
	0	272	18
Jc2_nrmtxt (SGD)	1	18	136
	0	278	12
JC2_cl_txt_img_emo_tag_pro_core_emp (Random Forest)	1	11	143

Table 15: Metrics of model performance for specific machine learning models on Cleveland justice data.

Model	**JC2_txt_tag_img**	**JC2_txt_img_pro_tag_emoji**	**Jc2_nrmtxt**	**JC2_cl_txt_img_emo_tag_pro_core_emp**
Measure	Value	Value	Value	Value
Sensitivity	0.9419	0.8874	0.8831	0.9226
Specificity	0.9723	0.9317	0.9379	0.9619
Precision	0.9481	0.8701	0.8831	0.9286
Negative Predictive Value	0.969	0.9414	0.9379	0.9586
False Positive Rate	0.0277	0.0683	0.0621	0.0381
False Discovery Rate	0.0519	0.1299	0.1169	0.0714
False Negative Rate	0.0581	0.1126	0.1169	0.0774
Accuracy	0.9617	0.9167	0.9189	0.9482
F1 Score	0.945	0.8787	0.8831	0.9256

Out of the Cleveland dataset, 246,802 tweets were classified as positive by at least one model. However, unlike the pattern seen in the Baltimore dataset, many tweets were classified as positive by only one of these four models.

This is driven, in large part, by the results of one of the classifiers (normalized text), which gave far more positive results than the others. This classifier produced over 80,000 positive results that were not identified by any of the others. Examination of a random sample of 50 of these tweets classified as positive by only this classifier found that most were false positives (44). False positives included tweets discussing issues of racism and protest that were not explicitly linked to the criminal justice system, and actions and events involving government and emergency response. These examples were all retweeted dozens of times:

- *RT @A: *black man is shot* it was his own fault, *girl is raped* she was asking for it *white boy shoots up a school* he's dis..*
- *RT @B: BUFFALO NY SNOW STORM: -75 inches -7 dead -State of emergency -National Guard deployed http://.. http://..*
- *RT @C: number of people who died for this protest=1 number of people who have died for this flag= 1.2 million #disrespect*
- *RT @D: Group 'Anonymous' claims responsibility for taking down City of Cleveland website http://.. http://..*

Others referenced sports. The remaining classifiers performed much more consistently, as was observed in the Baltimore dataset.

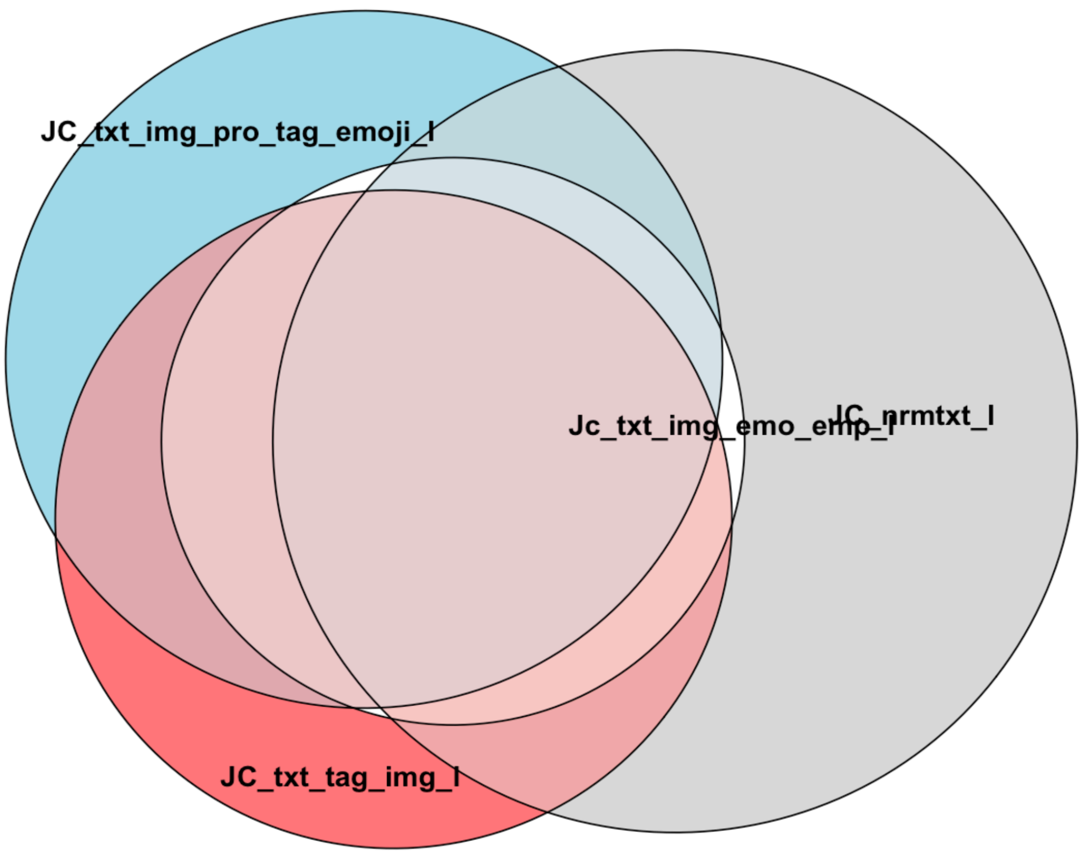


Figure 7: Overlap in agreement across Cleveland justice classifiers. With the exception of the normalized text classifier, most tweets are classified positive by the remaining classifiers.

As was done in the Baltimore analysis, ensemble models combing the results of these individual classifiers were explored. Setting a threshold of 2 or more models in agreement produces a set of 92,780 positive justice tweets for Cleveland. These tweets were created by 10,661unique users who tweeted about justice at any point during the timeframe. This method at this threshold improves performance in justice tweet classification.

Table 16: Performance of ensemble model for Cleveland justice classification (binary classification of justice/not justice, n=2271).

Model	**Cleveland**
Measure	Value
Sensitivity	0.964
Specificity	0.987
Precision	0.971
Negative Predictive Value	0.984
False Positive Rate	0.013
False Discovery Rate	0.029
False Negative Rate	0.036
Accuracy	0.980
F1 Score	0.968

Reviewing a random sample of 50 tweets classified as positive from Cleveland by this classifier, the vast majority are correct. Several are difficult to interpret due to lack of context or may be referencing events humorously or sarcastically:

- *RT @A: This is like when a girlfriend takes 30 minutes to tell you she wants to break up with you. #Ferguson*
- *I wonder how much #revcom-dot-us' web traffic has increased since plastering their URL on #Ferguson protest signs #tacky #exploitative*
- *Calvin's soteriology: efficient cause, God's free love; material cause, Christ; instrumental cause, faith; final cause, justice.*
- *So hot! Call the police and the fire man...*

3.7 Simpler models for tweet classification for justice

While the tweet classification models considered in this study used a large number of features, including several—such as image classification features—that require additional effort to generate, simpler models based on hashtag features[5] and a small set of language features also perform reasonably well against the labeled data used in this study.

This is illustrated first by considering several simpler models trained and tested using Baltimore data. Considering models incorporating only the Baltimore justice hashtag class feature, only the Freddie Gray hashtag class feature, all the Baltimore hashtag class features, a small set of language features[6], and a combination of language features with all Baltimore hashtag classes, all models achieve overall accuracy ranging from 77% to 88%. Confusion matrices and additional measures of model performance are supplied in the following tables.[7] Models considered were computed with the rattle package (v5.4.0; Williams, 2011), using Decision Tree (rpart), Extreme Boost (xgb), Random Forest (rf), and Logistic Regression (glm). Model performance was extremely similar or identical across these models for a specific set of features. For the cases in which the confusion matrix for multiple models was identical, only one is reported in the table below. In cases in which one model did slightly outperform others, based on accuracy and area under the ROC curve, that model is reported.

---

[5] Each of these hashtag features, such as the justice hashtag feature, is based on the set of relevant hashtags for that class, as described earlier. For example, there are dozens of distinct justice hashtags. Each tweet was scored for how many justice hashtags, and hashtags for the other categories, that it contained. See Chapter 3, *Hashtags and hashtag categories*, for more details on the hashtags in each class.

[6] These three language features were identified using the JRip algorithm in weka on the annotated Baltimore dataset, a dataset with over seven thousand language features, as well as hashtag and image classification features. The language feature variable names produced by the Cobalt system, such as "cobaltlang.nltk_normalized_text____dd6d2dcc67", are not human interpretable. See 3.3.1, Text processing, for more details on text processing. See 36, Tweet classification for justice, for more details on developing a ground truth dataset for tweet classification.

[7] Sensitivity (TPR = TP / (TP + FN)), Specificity (SPC = TN / (FP + TN)), Precision (PPV = TP / (TP + FP)), Negative Predictive Value (NPV = TN / (TN + FN)), False Positive Rate (FPR = FP / (FP + TN)), False Discovery Rate (FDR = FP / (FP + TP)), False Negative Rate (FNR = FN / (FN + TP)), Accuracy (ACC = (TP + TN) / (P + N)), F1 Score (F1 = 2TP / (2TP + FP + FN))

Table 17: Confusion matrices for best performing machine learning models using hashtag or hashtag and language features on Baltimore justice data (binary classification of justice/not justice, n=2462, 70/30 split)

Model	Confusion Matrix		
		Predicted	
	Actual	0	1
Justice hashtags (Decision Tree)	0	544	6
	1	157	32
Freddie Gray hashtags (Decision Tree)	0	530	20
	1	97	92
Baltimore hashtag classes (justice, Freddie Gray, riot, general) (Random Forest)	0	527	23
	1	86	103
Language features (Decision Tree)	0	535	15
	1	134	55
Baltimore hashtag classes and top language features (Random Forest)	0	514	36
	1	50	139

Table 18: Metrics of model performance for machine learning models using hashtag, language, or hashtag and language features on Baltimore test data

Model	Justice hashtags only	Freddie Gray hashtags only	All Baltimore hashtag classes	Language features	Hashtag and language features
Measure	Value	Value	Value	Value	Value
Sensitivity	0.1693	0.4868	0.545	0.291	0.7354
Specificity	0.9891	0.9636	0.9582	0.9727	0.9345
Precision	0.8421	0.8214	0.8175	0.7857	0.7943
Negative Predictive Value	0.776	0.8453	0.8597	0.7997	0.9113
False Positive Rate	0.0109	0.0364	0.0418	0.0273	0.0655
False Discovery Rate	0.1579	0.1786	0.1825	0.2143	0.2057
False Negative Rate	0.8307	0.5132	0.455	0.709	0.2646
Accuracy	0.7794	0.8417	0.8525	0.7984	0.8836
F1 Score	0.2819	0.6113	0.654	0.4247	0.7637

As this study is concerned with not just detecting justice tweets but also building predictive models of those users' engagement over time, performance on the positive class needs to be considered. Considering sensitivity and F1score, the justice hashtag model performs the most poorly. The Freddie Gray hashtag model, and the model incorporating multiple hashtag classes, both outperform this model for these metrics. The language features model also performs better than the justice hashtag model on these metrics. The model incorporating all hashtag and language features exceeds the performance of the other models, with 74% sensitivity and an F1 score of 76%.

For Cleveland, there are just two hashtag class features available: justice and Tamir Rice. Models were created incorporating only the Cleveland justice hashtag class feature, only the Tamir Rice hashtag class feature, both hashtag class features, and the small set of prior language features used for Baltimore. As the galvanizing event in Cleveland differed in multiple respects from the event in Baltimore, there could potentially be different language features that were important. For this reason, analogously to the approach used for Baltimore, another rules-based algorithm, the Decision Tree, was applied to the Cleveland training and testing data with the full set of Cleveland language features. This identified three additional language features. Models using the combined set of the prior Baltimore language features and the three additional language features identified from the Cleveland data, and one with a combination of these language features with both Cleveland hashtag classes were also created. As was done for the Baltimore case, models were computed with rattle, using Decision Tree (rpart), Extreme Boost (xgb), Random Forest (rf), and Logistic Regression (glm). Model performance was extremely similar or identical across these models for a specific set of features. Again, as with Baltimore, for the cases in which the confusion matrix for multiple models was identical, only one is reported in the table below.

For Cleveland, the models shown in the following table achieve high overall accuracy ranging from 74% to greater than 90%. Confusion matrices and additional measures of model performance are supplied in the following tables.

Table 19: Confusion matrices for best machine learning models using hashtag, language, or hashtag and language features on Cleveland justice data (binary classification of justice/not justice, n=2271, 70/30 split):

Model	**Confusion Matrix**		
	Actual	**Predicted** 0	1
	0	463	3
Justice hashtags (Decision Tree)	1	166	34
	0	466	0
Tamir Rice hashtags (Decision Tree)	1	169	31
	0	463	3
Justice and Tamir Rice hashtags (Decision Tree)	1	145	55
	0	459	7
Prior language features (Decision Tree)	1	109	91
	0	444	22
Combined language features (Logistic regression)	1	46	154
	0	444	22
Combined language features and hashtags (Boost)	1	19	181

Table 20: Metrics of model performance for machine learning models using hashtag or hashtag and language features on Cleveland test data

Model	**Justice hashtags**	**Tamir Rice hashtags**	**Justice and Tamir Rice hashtags**	**Prior language**	**Combined language**	**Combined language and hashtags**
Measure	Value	Value	Value	Value	Value	Value
Sensitivity	0.17	0.155	0.275	0.455	0.875	0.905
Specificity	0.9936	1	0.9391	0.985	0.9061	0.9528
Precision	0.9189	1	0.6471	0.9286	0.77	0.8916
Negative Predictive Value	0.7361	0.7339	0.7615	0.8081	0.9528	0.959
False Positive Rate	0.0064	0	0.0609	0.015	0.0939	0.0472
False Discovery Rate	0.0811	0	0.3529	0.0714	0.23	0.1084
False Negative Rate	0.83	0.845	0.725	0.545	0.125	0.095
Accuracy	0.7462	0.7462	0.7475	0.8258	0.8979	0.9384
F1 Score	0.2869	0.2684	0.386	0.6107	0.8191	0.8983

Considering sensitivity and F1 scores, the Tamir Rice hashtag model performs the most poorly, followed by the justice hashtag model. The model incorporating both justice and Tamir Rice hashtag classes is also weak. The prior language features model exceeds the hashtag models, and the final two models improve considerably on the simplest models. They achieve sensitivity of 86% to 95%, and an F1 score from 82% to 89%.

Overall, this exploration of simpler models to classify justice-related tweets suggests that models that just incorporate hashtags and language features may be sufficient, depending on the goal of the research. The discrepancies in performance between these sets of Baltimore and Cleveland models may suggest that aspects of the events and community response shaped the discourse differently in these two cases. Given the goals and focus of this study—to identify factors associated with continued engagement in justice-related conversations in the weeks and months after a galvanizing event —I use the more complex model described in the previous section to build my corpus of justice-related tweets that are then used in models to predict engagement.

3.8 Distinguishing and characterizing highly engaged vs. cursory actors in justice discourse

For those actors who participated in justice discourse, but had not been active prior to the death of Freddie Gray or Tamir Rice, respectively, I explored whether we can distinguish those who remain in the community (my dependent variable) from those whose participation is fleeting. This line of inquiry addresses the question: How are those who become more engaged and remain engaged over time different from those who do not, based on their social media interactions and connections, identity presentation, expressions of emotion, and effort? Improving knowledge of who becomes committed to justice can help us better understand social movements and other efforts to spur social change, and how these processes unfold in digital environments. It could help motivate designs for tools and technology to mobilize and galvanize political participation through social media platforms.

A threshold was developed in consideration of the justice behavior of the Baltimore actors who were active early (during the T1 baseline, prior to the death of Freddie Gray) with respect to justice. As there are high-activity outliers among these users, median activity is considered. Most of these baseline T1 actors tweeted about justice once during the week of T1 (median). During T3, post-event, all T1 baseline actors tweeted four times (median), and T1 actors who continued tweeting about justice in T3 tweeted five times (median) over that one-month time period.

It's also important to consider that actors who were already engaged in justice might be more active than newcomers. The median activity for these new Baltimore users who began tweeting about justice in T2 and continued in T3 is two, lower than that of the baseline actors. This suggests a threshold of engagement for the Baltimore dataset of four or more tweets is appropriate, as it aligns with T1 actors' levels of activity and does not dramatically exceed median activity of the newcomer population (two more tweets, or double their median). At this threshold, roughly 20% of new T2 actors remain engaged in T3.

A similar process was applied to Cleveland. Median levels of activity for baseline Cleveland actors during T1 were the same as for Baltimore, at one tweet. In Cleveland, T3 spans a longer period, due to the slower pace of events in the community[8], as well as the lower amount of justice activity. During T3, all T1 baseline actors tweeted three times (median), and T1 actors who were active in T3 tweeted six times (median), a similar but slightly wider range than

---

[8] In Baltimore, the medical examiner's report declaring a homicide was released within about 10 days and the State's Attorney began proceedings against the officers immediately. In Cleveland, it took over a month for homicide to be declared, and investigation of the officers began the following year in another jurisdiction.

Baltimore. The median activity for the new Cleveland users who began tweeting about justice in T2 and continued in T3 is the same as Baltimore, at two. Setting a threshold of four tweets for Cleveland also aligns with median tweeting rates for Cleveland T1 baseline actors in T3, does not dramatically exceed median activity of the Cleveland newcomer population (two more tweets, or double their median), and also leads to 20% of new T2 actors remaining engaged in T3. As this is just one of a number of possible approaches for establishing a threshold, predictive models were also run with both higher and lower thresholds for engagement, with comparable results. See Appendix D for more details. Future work could explore alternative, more complex methods of defining ongoing engagement.

For the highly engaged actors, this work examined how and whether *identity*, *emotion*, *effort*, and *social embeddedness* predict likelihood that an individual stays engaged by continuing to tweet about justice after the main event passes (i.e., during T3). This necessitated examining the available digital trace data for variables that could reasonably represent these factors. For identity, these included aspects of the user profile, where users name and describe themselves, and select images to represent themselves. For emotion, emotion emoji and Empath variables for emotion were available. Of course, alternate processing pipelines could provide different options for variables representing these factors. For effort, potential digital traces include actions that enrich a tweet, while suggesting increased cognitive (or physical) demand. Content enhancement via images posted and shared on social media has been associated with semiotic significance, actively shaping the discursive construction of an event, and images can be strategically employed by users in elaborating discourse, and to overcome platform-imposed constraints on utterances (Morin et al., 2019). Different images added to identical text can be used to inflect an event or topic's framing. Similarly, an identical image can be added to distinct messages in the context of social debates. The inevitable context collapse with Twitter, leading to users communicating with an "imagined" audience (D. boyd et al., 2010) challenges users in terms of contextual assumptions. Employing hashtags is a mechanism by which users can explicitly communicate otherwise implicit aspects of meaning, guiding a reader's inferential processes while maintaining style and tone (Scott, 2015). Table 21 provides example variables for each of these four dimensions.

Additionally, while retweets do not incur the effort of typing out a message, for the user who retweets, this action reflects the cognitive effort and decision to amplify or spread content to new audiences, to inform specific audiences, to publicly agree, to validate other's perceptions and expressions, to express loyalty or alignment to an individual or a cause, or to integrate or support new or less visible individuals and content (D. boyd et al., 2010). Further, work on Internet-based political mobilization suggests a "ladder of participation" in which lower intensity online activities can lead to more intensive forms of engagement (Cantijoch et al., 2016) . For these reasons, retweets are included, as well as original tweets and replies.

Table 21: Summary of dimensions and associated variables. These variables will be measured at T2, the time period of the riots, for users who tweeted about justice, and should help predict if an actor will remain engaged in justice discourse at T3, after the riots ended and when general interest and news media coverage has ebbed.

Dimensions	**Example variables**
Identity/identification	Changed user profile image Changed screen name Changed user description Use of forms of pronouns such as "we"
Emotion	Use of emotion terms Use of emotive emoji
Effort	Freddie Gray (Tamir Rice) hashtags used Justice_hashtags used Riot_hashtags used Police images in tweets Peaceful protest images in tweets

	Riot images in tweets
Social embeddedness	In-degree centrality Out-degree centrality Closeness centrality Betweenness centrality

3.9 Determining actor-level indicators for predicting sustained justice engagement

To move from the curated tweet justice datasets for Baltimore and Cleveland to data suitable for making actor-level predictions about continued engagement in justice discourse on social media, it was necessary to both compute new variables about the actors (users who tweeted about justice), and to appropriately aggregate tweet-level indicators into actor level indicators that may reasonably correspond to dimensions of interest (identity, emotion, social embeddedness, and effort) and other factors of interest. Social network analysis provided a window into actor positions in the larger justice discourse network within each city, enabling a quantification of social embeddedness.

3.9.1 Social network analysis of tweet-based justice networks

Huberman et al. (2008) point out weaknesses of using networks based on following behaviors when studying the propagation of ideas or the formation of social ties, as opposed to studying networks derived from interactions based on message content. They characterize the latter as social networks that "matter" (Huberman et al., 2008). Since I am interested in networks reflecting justice-related discourse, the focus of this research is on tweet content. For this analysis, I constructed networks of actors (Twitter users) connected by ties that reflect their interaction with other users. Within tweets, the types of interactions extracted are described in Table 22.

Table 22: Tie types used in social network analysis

Type	Description
Reply	Tweet that is explicitly replying to another
Direct mention	Explicitly naming another in the content of a tweet, without having directed the tweet at that other, comparable to talking about another in public
Retweet (RT)	Rebroadcasting the tweet of another, comparable to quoting another in public
Self	A tweet that does not mention another user (can be represented as an isolate or a self-loop in a network)

Simply creating a network of user mentions conflates behaviors that are sociologically and semantically distinct, and may blur relevant distinctions between highly engaged actors and cursory actors. These interactions were extracted from the set of relevant tweets. This goes beyond a network of users who sent or were mentioned in justice-related tweets to more richly characterize the nature of the engagement in the network. Pointillist graph visualization software[9] was used to extract all user mentions and determine the interaction type relative to the tweet sender from the tweet JSON.

After appropriate networks were constructed based on justice tweets from T2, social network analysis metrics were calculated for all actors who appeared in the networks. These included multiple actor-level metrics. Actor-level measures provide an understanding of how individual actors—users who have tweeted about justice— "fit" into the larger network structure. Differences in actor level measures may correlate with whether an actor is influential in the community, or whether an actor becomes more highly engaged.

Among the most frequently used actor level measures are those that attempt to assess an actor's centrality or prestige. *Degree centrality* assesses how prominent, or "central" an actor is, based on how many ties the actor has to others in the network (Freeman, 1979). Indegree

[9] Cohen, J. 2021. Pointillist [graph analysis and visualization program], Version 2.5, Applied Physics Laboratory, Johns Hopkins University.

captures how many ties come in to an actor, while outdegree measures how many ties originate with the actor and connect with others.

Subgroup-level measures provide a sense of how fragmented or clustered the network structure is. A network with many distinct subgroups or clusters may hold substantially different views or positions on issues, or otherwise fail to cohere. For example, the existence of major, clearly distinct subgroups within the justice network could reflect a partitioning into groups that perceive a lack of procedural justice, as opposed to those who are satisfied with justice in their cities. For this analysis, I determined the distinct components within the justice discourse network, and whether each actor was a member of the main component. This network approach was repeated on the justice Twitter data from Cleveland.

Network level measures can provide a high-level structural perspective, on overall patterns of size and connectivity. A weakly connected network, for example a network with many disconnected pieces (components), or one with very low density and long path lengths, may be less likely to add many high engagement actors. A set of social network measures examined in this analysis is provided below. Highly engaged actors should be distinguishable through their network and positional metrics, in addition to their frequency of tweet behaviors.

Table 23: Sample social network metrics considered for assessing social embeddedness of actors. These metrics helps characterize the prominence, or centrality, of actors as well as whether they are situated within the main component or a smaller component of the network.

Type	Measure	Description	Purpose
Actor	In-degree centrality	Number of others who form ties with actor	Identify actors who are prominent in network because many others reference or interact with them
	Out-degree centrality	Number of others that actor forms ties with	Identify actors who are prolific in connecting with others
	Betweenness centrality	Number of shortest paths that flow through actor	Identify actors who are potential conduits for information flow or influence
	Closeness centrality	How "close" as a sum of path lengths an actor is to others	Identify actors in closer proximity to many others as opposed to peripheral positions

Subgroup	Components	Segments of the network that are disconnected from each other. The main component (MC) is the largest component in the network.	Provides another perspective on amount of connectedness or fragmentation of the network

3.9.2 Empath variables for predictive modeling

For each tweet, Empath computed the percentage of words in the tweet that corresponded to an Empath category, ranging from 0 to 1. For many tweets, most Empath categories score 0.0. The Empath variables scores for positive justice tweets were reviewed to identify a subset of variables which appeared with reasonable frequency in justice tweets and which pertained to emotion dimensions or had topical relevance. These Empath variables were included in the predictive models for user engagement.

Reviewing the Empath variables that were ranked in the top 25 for count of tweets or for mean value produced a set of 45 Empath variables, of which 25 were assessed to be potentially relevant, as they reflected emotions or were pertinent to criminal justice or the events.

Table 24: Empath variables retained for predictive models

Empath Variable Name		
aggression	order	suffering
fight	shame	stealing
positive_emotion	exasperation	negative_emotion
kill	communication	pain
death	crime	listen
sympathy	prison	law
achievement	dispute	violence
religion	speaking	weapon
horror		

3.9.3 Creating features for predictive models of user engagement

For each user who had tweeted about justice, a representation of their justice-related Twitter activity was constructed. This summarized information about their behaviors, activity, and network position during T2. These variables were organized around the dimensions of identity, emotion, effort, and social embeddedness.

Table 25: Variables used in predicative models for user engagement after a galvanizing event. These variables were included in logistic regression models for predicting sustained engagement with justice discourse in T3, post event. For each variable, the aggregation method, and relevant dimensions are shown.

Dimension	Variable	Type	Aggregation method
Effort	FG_tags	integer	count of tweets containing relevant hashtag
	TR_tags	integer	count of tweets containing relevant hashtag
	J_tags	integer	count of tweets containing relevant hashtag
	R_tags	integer	count of tweets containing relevant hashtag
	person_emoji	integer	count of tweets containing relevant emoji
	pic_people	integer	count of tweets with image classifier score > .5
	pic_riot	integer	count of tweets with image classifier score > .5
	pic_protest	integer	count of tweets with image classifier score > .5
	pic_parade	integer	count of tweets with image classifier score > .5
	pic_police	integer	count of tweets with image classifier score > .5
Emotion	pos_emoji	integer	count of tweets containing relevant emoji
	neg_emoji	integer	count of tweets containing relevant emoji
	aggr	integer	count of tweets with Empath value >0
	fight	integer	count of tweets with Empath value >0
	posemo	integer	count of tweets with Empath value >0
	kill	integer	count of tweets with Empath value >0
	death	integer	count of tweets with Empath value >0
	symp	integer	count of tweets with Empath value >0
	horror	integer	count of tweets with Empath value >0
	shame	integer	count of tweets with Empath value >0
	exasp	integer	count of tweets with Empath value >0
	suffer	integer	count of tweets with Empath value >0
	negemo	integer	count of tweets with Empath value >0
	pain	integer	count of tweets with Empath value >0
General	count	integer	count of justice tweets

	activity	integer	tweets in T2
	favourites_count	integer	maximum tweet value
	followers_count	integer	maximum tweet value
	friends_count	integer	maximum tweet value
	statuses_count	integer	maximum tweet value
Identity	prof_image_change	binary	1> profile images <4
	descrip_change	binary	1> profile descriptions <4
	i	integer	count of tweets with form of pronoun
	we	integer	count of tweets with form of pronoun
	you	integer	count of tweets with form of pronoun
	they	integer	count of tweets with form of pronoun
Social embeddedness	Betweenness	0.0-1.0	network centrality score
	Closeness	integer	network centrality score
	Degree	integer	network centrality score
	InTieCount_DirectMention	integer	network centrality score
	InTieCount_ReplyTo	integer	network centrality score
	InTieCount_Retweet	integer	network centrality score
	OutTieCount_DirectMention	integer	network centrality score
	OutTieCount_ReplyTo	integer	network centrality score
	OutTieCount_Retweet	integer	network centrality score
	inMC	binary	network metric
Topical	achiev	integer	count of tweets with Empath value >0
	religion	integer	count of tweets with Empath value >0
	order	integer	count of tweets with Empath value >0
	comms	integer	count of tweets with Empath value >0
	crime	integer	count of tweets with Empath value >0
	prison	integer	count of tweets with Empath value >0
	dispute	integer	count of tweets with Empath value >0
	speak	integer	count of tweets with Empath value >0
	steal	integer	count of tweets with Empath value >0
	listen	integer	count of tweets with Empath value >0
	law	integer	count of tweets with Empath value >0
	violence	integer	count of tweets with Empath value >0
	weapon	integer	count of tweets with Empath value >0

For completeness, several other variables characterizing user engagement with Twitter were considered, such as friends count and followers count. In addition to these general variables, a subset of Empath variables that appear topically relevant (pertaining to the law, crime, weapons, violence etc.) were also considered.

Examination of the distribution of values for the numeric variables showed most did not have a normal distribution, which is not uncommon for social media or social network-based data. To reduce skew and decrease correlations between variables, those variables were transformed by taking their square root, and using that value in the logistic regression. This also reduced the variance influence factor (VIF) for most variables to between one and three. As the person and emoji variables, as well as the image variables capture overlapping classes, they continue to have higher VIFs. As Closeness was not skewed, its original values were used in logistic regressions.

I applied logistic regression to predict likelihood that an individual stays engaged, and to determine whether dimensions and their variables were significant in predicting this outcome. More detailed discussion of the logistic regression models can be found in Chapter 4.

3.10 Summary of approach and methods

This chapter has detailed the approach and methods used to identify justice discourse from two large Twitter datasets collected from Baltimore and Cleveland around the time of two galvanizing events that lead to the deaths of two community members, a young African American man and an African American child, at the hands of police. In both events, these deaths triggered widespread outrage, local protests, and calls for change to policing and the larger criminal justice system.

Tweets from Baltimore and Cleveland were processed and enriched. Machine learning models were developed to classify each tweet as justice-related or not. The justice tweet dataset for each city, comprising a small percentage of the original tweets, was used to construct actor-level representations of all users who tweeted about justice. These representations summarized user behavior and activity for variables that were intended to capture *identity, emotion, effort,* and *social embeddedness*. These dimensions were hypothesized to predict ongoing engagement in a digital environment, as they have for face-to-face collective action and social movements.

For completeness, additional variables reflecting general platform properties of the users and use of certain categories of topical language were also considered. Logistic regression was performed on those actors who joined their communities' social media discourse about justice during T2 (the peak of the event), to predict which actors would sustain engagement during T3, post-peak, and to examine which dimensions and variables were significant. These results are discussed in Chapter 4.

Chapter 4: Analysis and findings

The goal of this dissertation is to explore expressions of justice on Twitter following galvanizing events and identify factors that predict continued engagement in discussions of justice on the platform in the weeks and months after those events. To do this, machine learning was applied to two large Twitter datasets from Baltimore and Cleveland from the timeframes surrounding the homicides of Freddie Gray and Tamir Rice to identify tweets relating to justice. These tweets were partitioned into a pre-event baseline of justice social media posts (T1), a second set that occurred during the peak of community response to these galvanizing killings (T2), and a post-event set (T3).

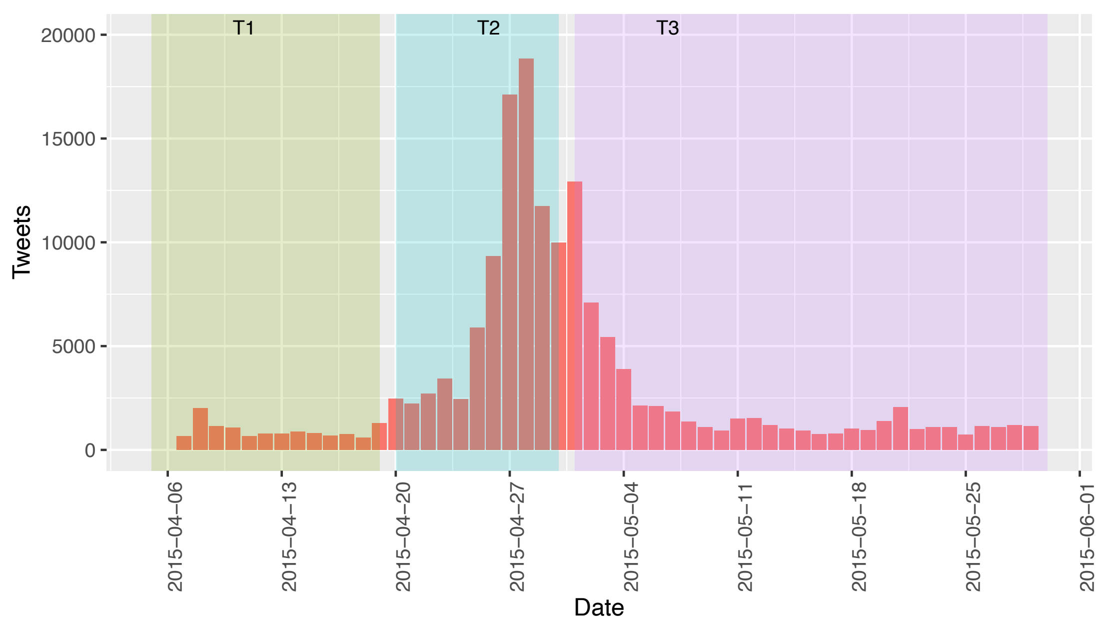


Figure 8: Justice tweets per day in Baltimore. From a lower baseline rate (T1), justice tweeting increases substantially in T2, then drops in T3, though activity remained elevated compared to the earlier T1 rate.

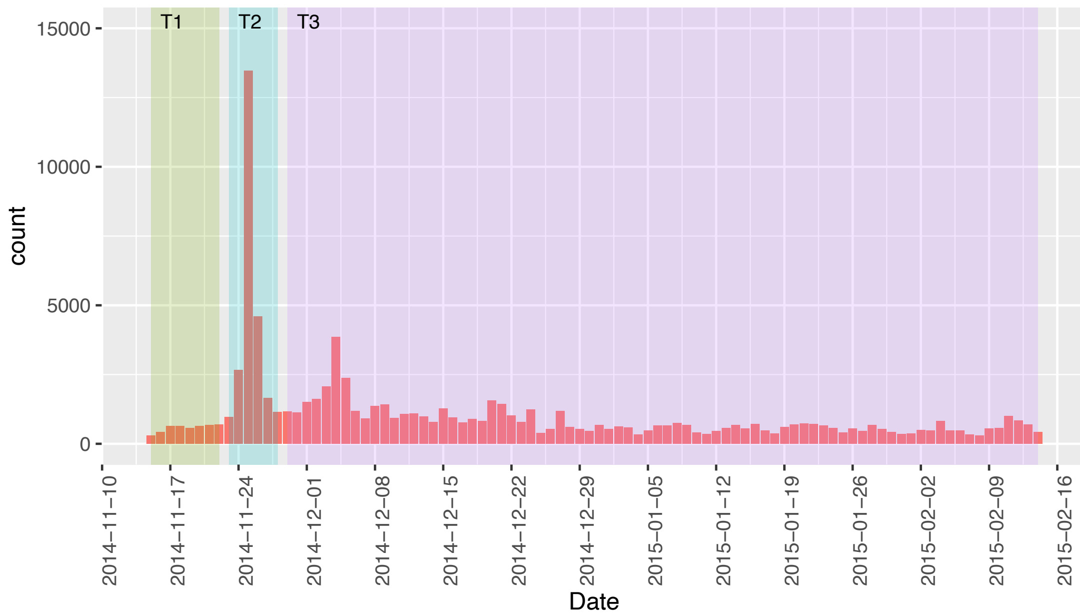


Figure 9: Justice tweeting activity in Cleveland. From a lower baseline at T1, activity spikes dramatically during the peak of the event (T2). A smaller increase persists through T3, the post-event time period.

Logistic regression models were created to predict which actors would remain engaged in justice discourse post-event (T3), after posting about justice in T2 initially. Variables used in these models are intended to reflect the dimensions of *identity, emotion, effort,* and *social embeddedness*, as well as general Twitter engagement and several topical variables, and were transformed to reduce skew as previously described in 3.9. Model findings from a threshold of four or more justice tweets for determining cursory vs. engaged participants at T3 are described in this chapter. Results from slightly higher or lower thresholds are presented in Appendix D.

In both communities, a sizable proportion of Twitter users posted about justice. In Baltimore, over one-quarter (26%, 15,881 users) were active at some point during the timeframe, while nearly one-fifth (18%, 10,661 users) tweeted about justice across the longer Cleveland timeframe. Both communities also saw a surge in participants with justice-related posts during T2. In Baltimore, the number nearly tripled from T1 to T2, growing from 3810 to 11,354 users.

The increase in Cleveland was equally striking, from 1560 to 4592 users. After both galvanizing events, a substantial percentage of the new justice discourse participants continued to stay engaged post event, during T3. For Cleveland, slightly more than 20% continued participating in justice discourse at or above the threshold (756 of 3673 during T3, while for Baltimore the ratio was just under 20% (2159 of 11,354).

4.1 Overview of engagement models

To facilitate describing the findings from these predictive models, overviews of models and model performance will be followed by more detailed descriptions of findings on dimensions and their respective significant variables. Tables 26 and 27 provide full details from the logistic regression models predicting sustained engagement in justice discussions for Baltimore and Cleveland, respectively. Further interpretation is presented in Chapter 5.

Table 26: Logistic regression model for sustained justice engagement in Baltimore, at a threshold of engagement of 4 or more tweets during T3. This model achieves a Pseudo R-Square (optimistic) of 0.643. (Sig. codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1)

	Estimate	Std. Error	z value	Pr(>|z|)	Sig
(Intercept)	-4.2374103	0.24231112	-17.487	<2.00E-16	***
count	0.65251762	0.20929845	3.118	0.001823	**
activity	0.10596712	0.01082948	9.785	<2.00E-16	***
FG_tags	0.03520108	0.05732522	0.614	0.539176	
J_tags	0.2537034	0.05732489	4.426	9.613E-06	***
R_tags	-0.2171266	0.07535989	-2.881	0.003962	**
pos_emoji	-0.1604837	0.16511932	-0.972	0.331088	
person_emoji	-0.0006315	0.14780347	-0.004	0.996591	
neg_emoji	-0.1514106	0.15570884	-0.972	0.330854	
i	-0.1048229	0.05912182	-1.773	0.076229	.
we	-0.0713659	0.0616824	-1.157	0.247277	
you	-0.2328811	0.06311916	-3.69	0.000225	***
they	-0.0539489	0.0659526	-0.818	0.41336	
aggr	-0.145767	0.11000557	-1.325	0.185142	
fight	-0.1332874	0.09979611	-1.336	0.181681	

posemo	-0.120142	0.0844048	-1.423	0.15462	
kill	0.28587717	0.11634828	2.457	0.014007	*
death	0.03484389	0.10483615	0.332	0.739614	
symp	0.3671562	0.22946869	1.6	0.109592	
achiev	0.19736869	0.12699339	1.554	0.120145	
religion	0.0287072	0.12756212	0.225	0.821944	
horror	0.31806581	0.14545259	2.187	0.028762	*
order_ave	-0.0453206	0.1075174	-0.422	0.673376	
shame	-0.0188119	0.11514704	-0.163	0.870225	
exasp	-0.5541589	0.27599319	-2.008	0.044657	*
comm	0.01198829	0.09706798	0.124	0.901708	
crime	-0.0724614	0.10085381	-0.718	0.472462	
prison	-0.0228475	0.09885309	-0.231	0.817217	
dispute	-0.0805043	0.07636334	-1.054	0.291779	
speak	0.07554115	0.09319466	0.811	0.41761	
suffer	-0.1848863	0.10939892	-1.69	0.091024	.
steal	0.16404902	0.10486298	1.564	0.117721	
negemo	-0.1907249	0.08488036	-2.247	0.024641	*
pain	-0.0675422	0.10810048	-0.625	0.532096	
listen	0.00333135	0.09361182	0.036	0.971612	
law	0.14968188	0.10292647	1.454	0.145874	
violence	0.22183442	0.10632927	2.086	0.036952	*
weapon	-0.172882	0.08569417	-2.017	0.043651	*
pic_people	0.12385487	0.36467967	0.34	0.734138	
pic_riot	0.11612451	0.3155549	0.368	0.712873	
pic_protest	-0.1866915	0.35581092	-0.525	0.599797	
pic_parade	0.57453832	0.27376993	2.099	0.035851	*
pic_police	-0.3205502	0.27747428	-1.155	0.247991	
Betweenness	-7.5284279	3.500253	-2.151	0.03149	*
Degree	-0.1845185	0.14011799	-1.317	0.187879	
InTieCount_DirectMention	0.09865282	0.08856309	1.114	0.26531	
InTieCount_ReplyTo	0.05861868	0.18482752	0.317	0.751127	
InTieCount_Retweet	0.1075513	0.07633706	1.409	0.158865	
OutTieCount_DirectMention	0.27137131	0.06764164	4.012	6.0233E-05	***
OutTieCount_ReplyTo	-0.2069508	0.11820384	-1.751	0.079981	.
OutTieCount_Retweet	0.07846848	0.08529258	0.92	0.357577	
favourites_count	-0.0039507	0.0014204	-2.781	0.005413	**
followers_count	-0.001087	0.00232696	-0.467	0.640409	

friends_count	0.0070974	0.00311472	2.279	0.022687	*
statuses_count	0.00335793	0.0006329	5.306	1.12E-07	***
Closeness_orig	0.00007237	0.00007444	0.972	0.330971	
TFC_prof_image_change	-0.2968033	0.15324746	-1.937	0.052775	.
TFC_descrip_change	-0.1069046	0.18140116	-0.589	0.555642	
TFC_inMC	-0.0980508	0.35342768	-0.277	0.781451	

Table 27: Logistic regression model for sustained justice engagement in Cleveland, at a threshold of engagement of 4 or more tweets during T3. This model achieves a Pseudo R-Square (optimistic) 0.719. (Sig. codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1)

	Estimate	Std. Error	z value	Pr(>|z|)	Sig
(Intercept)	-4.8009	0.6005	-7.995	1.29E-15	***
count	-1.6333	0.4149	-3.936	8.27E-05	***
activity.x	0.1362	0.0421	3.238	0.0012	**
TR_tags	-0.1855	0.1878	-0.988	0.3232	
J_tags	-0.5394	0.1308	-4.124	3.73E-05	***
pos_emoji	-0.1075	0.6564	-0.164	0.8699	
person_emoji	-0.4911	0.7184	-0.684	0.4943	
neg_emoji	-0.4838	0.7098	-0.682	0.4954	
i	0.0357	0.1361	0.263	0.7929	
we	-0.0210	0.1577	-0.133	0.8939	
you	-0.2268	0.1512	-1.5	0.1336	
they	-0.0195	0.1750	-0.112	0.9112	
aggr	0.1453	0.2703	0.537	0.5909	
fight	0.3175	0.2356	1.348	0.1777	
posemo	-0.2651	0.2015	-1.315	0.1884	
kill	-0.1207	0.2698	-0.447	0.6547	
death	0.0771	0.2552	0.302	0.7626	
symp	0.7056	0.5373	1.313	0.1891	
achiev	-0.9566	0.3887	-2.461	0.0139	*
religion	-0.1341	0.3991	-0.336	0.7370	
horror	-0.6624	0.4408	-1.503	0.1329	
order_ave	0.0220	0.2852	0.077	0.9386	
shame	0.3604	0.2807	1.284	0.1992	
exasp	0.8611	0.5825	1.478	0.1393	
comm	-0.3056	0.2435	-1.255	0.2095	
crime	0.0045	0.2092	0.022	0.9828	
prison	0.0341	0.2466	0.138	0.8902	
dispute	0.1449	0.1694	0.856	0.3922	
speak	0.3881	0.2001	1.939	0.0525	.
suffer	-0.0248	0.2640	-0.094	0.9253	
steal	0.1135	0.2320	0.489	0.6247	
negemo	0.0070	0.1676	0.041	0.9669	
pain	-0.1780	0.2699	-0.659	0.5097	
listen	-0.0262	0.2329	-0.113	0.9103	
law	0.0218	0.2502	0.087	0.9304	
violence	0.2179	0.2593	0.84	0.4007	
weapon	-0.1484	0.1721	-0.863	0.3884	
pic_people	0.2657	0.5984	0.444	0.6570	
pic_riot	0.4957	0.6472	0.766	0.4438	

pic_protest	0.0214	0.5674	0.038	0.9699	
pic_parade	-1.3079	0.5946	-2.2	0.0278	*
pic_police	1.1761	0.4842	2.429	0.0151	*
Betweenness	7.8119	4.2905	1.821	0.0687	.
Degree	1.1680	0.3631	3.217	0.0013	**
InTieCount_DirectMention	-0.1071	0.1952	-0.548	0.5835	
InTieCount_ReplyTo	0.3792	0.3890	0.975	0.3297	
InTieCount_Retweet	0.3324	0.2175	1.529	0.1264	
OutTieCount_DirectMention	0.3962	0.1425	2.78	0.0054	**
OutTieCount_ReplyTo	0.5932	0.2465	2.407	0.0161	*
OutTieCount_Retweet	0.9759	0.2016	4.84	0.0000	***
favourites_count	-0.0078	0.0028	-2.8	0.0051	**
followers_count	0.0042	0.0022	1.902	0.0572	.
friends_count	0.0027	0.0050	0.548	0.5835	
statuses_count	0.0050	0.0012	4.119	0.0000	***
Closeness_orig	0.0004	0.0001	3.137	0.0017	**
TFC_prof_image_change	0.0744	0.6075	0.123	0.9025	
TFC_descrip_change	-0.1258	0.6671	-0.189	0.8505	
TFC_inMC	-2.4989	0.7263	-3.441	0.0006	***

Both models outperform the null hypothesis substantially. They classify the cursory class (not engaged) with ~97% accuracy for both Baltimore and Cleveland. Correct classification of the engaged actors is more difficult, with 48% and 59% accuracy, driving overall error to 12-13% for Baltimore and Cleveland. Receiver operating curves for both models are shown below.

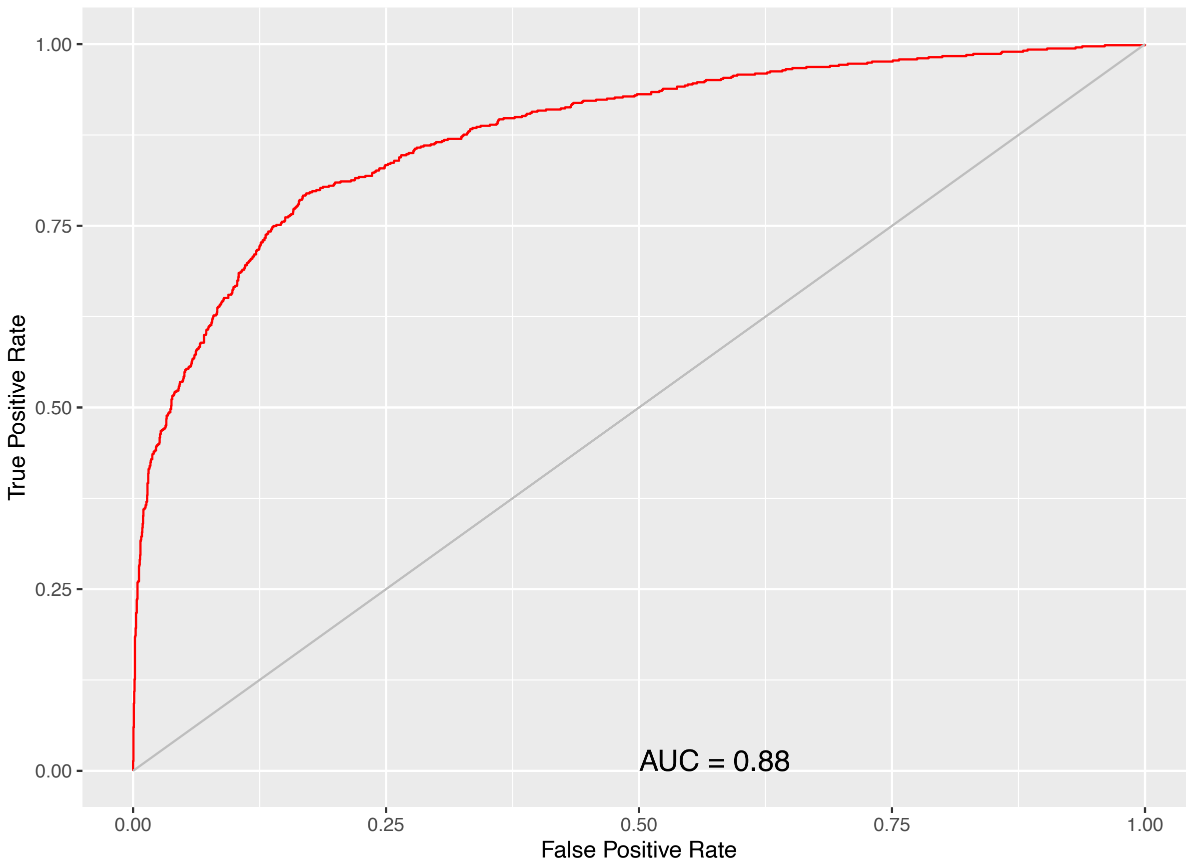


Figure 10a: Receiver Operating Curve (ROC) for Baltimore model. The area under the curve is slightly higher for Cleveland than Baltimore.

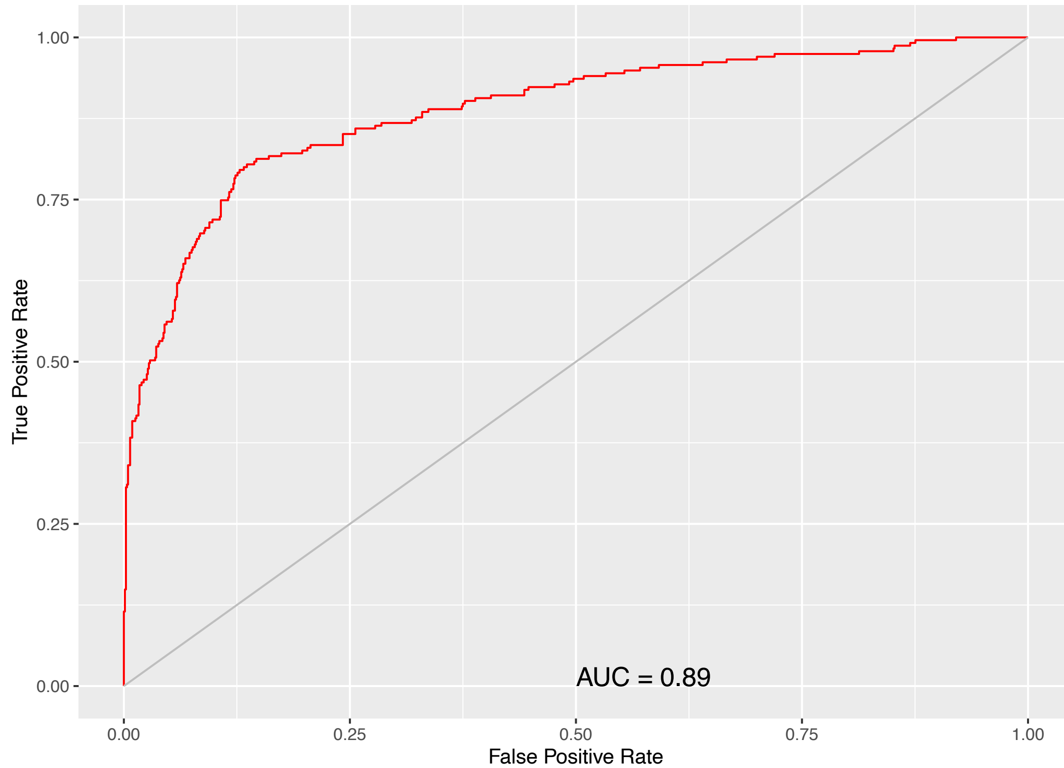


Figure 11b: Receiver Operating Curve (ROC) for Cleveland model. The area under the curve is slightly higher for Cleveland than Baltimore.

In the following sections, I will break down significant and non-significant findings across the main categories of predictors: identity, emotion, effort, and social embeddedness, as well as a discussion of the topical and general predictor variables. I will also briefly review the consistency of engagement models at the different thresholds.

4.2 Identity

Social network sites such as Twitter offer their users the opportunity to construct a public representation of self through various elements of their profile, including the user's screen name or handle, profile image, and profile description. This "enactment of a digital identity" (danah boyd & Crawford, 2012) affords individuals who become affiliated with a social movement, cause, or issue a means to reflect that identity in their profile. For example, following the death of Tamir Rice or Freddie Gray, a person may have added the Black Lives Matter hashtag to their profile, or changed their profile image to reference the events. This work explored whether changes in these components of a user's profile during the timeframe around the two events would predict ongoing engagement with justice discourse.

In both Baltimore and Cleveland, variables capturing changes to elements of the user profile were not significant predictors of continued engagement in justice discourse. For individuals from these communities, at this juncture in time, the number of individuals who exhibited these behaviors was quite small.[10] Of the Baltimore T2 users in the model, only a few hundred changed their profile image, description, or name; in Cleveland, only a few dozen T2 users did so. In the Baltimore dataset, just 17 of the newly active T2 users changed the screen names, while only one did so in Cleveland. Use of pronouns that reflect shared identity (we), as opposed to focus on others (you, they) was also considered as a potential predictor of sustained engagement. While this variable was not significant in either of these models at the threshold, use of the *you* variable was negatively associated for the Baltimore users. Overall, the identity dimension, as captured in these logistic regression models, was not exceptionally useful for predicting continued engagement.

<u>4.3 Emotion</u>

A subset of Empath variables was used to quantify expression of emotion in the text of tweets for both datasets.[11] This was aggregated to capture expression of emotion at the individual

[10] Upon recommendation of the dissertation committee, these variables were operationalized to specify users who only had two or three changes during the time period. Users who changed their profile image dozens of times, for example, were unlikely to be reflecting identity change relating to the issue of justice.
[11] The full list of Empath variables is presented in Appendix C, while Table 24 summarizes those that were considered in this study.

level. A few variables associated with the emotion dimension were significant in Baltimore. For example, mentions of terms associated with horror and killing elevated the likelihood of engagement, while use of terms associated with suffering did not. While those that provided more general reflections of emotion through emojis (pos_emoji, neg_emoji) were not significant, negemo was significant.

The pattern varied somewhat when looking at the Cleveland data. General reflections of emotion through emoji and Empath emotion variable posemo and negemo were not significant, nor was horror (although it was significant at a lower threshold of three or more tweets). The remaining emotion variables were not significant.

4.4 Effort

For actors who are newly engaged in justice discourse in T2 in Cleveland, incorporating and sharing images containing police in posts increased the probability of continued engagement in T3. Justice hashtags, another indicator of effort, were mixed in effect, being a positive predictor in Baltimore while slightly negative in Cleveland. Hashtags referencing the specific victims of police violence in these two events (Freddie Gray and Tamir Rice), did not reach significance, while use of riot hashtags in Baltimore was negatively predictive.

4.5 Social Embeddedness

The social network variables used to capture social embeddedness reflect each actor's position within the network structure, as well as choices made by the actor toward others (OutDegree) in a justice tweet, or made by other others toward the actor (InDegree) in their justice tweets. A description of each network is presented to provide context for interpreting this dimension.

In T2, the network constructed from Baltimore justice tweets includes 25,604 total actors. Of these, over 70% are in the main component. The remaining components are very small, and about 20% of actors are isolates, unconnected form all others in the graph. Of the 2199 actors engaged during T3, the vast majority are connected to others, with about 80% in the main component and just 9% isolates. The figure below shows the Baltimore network at T2, highlighting the positions of the engaged actors.

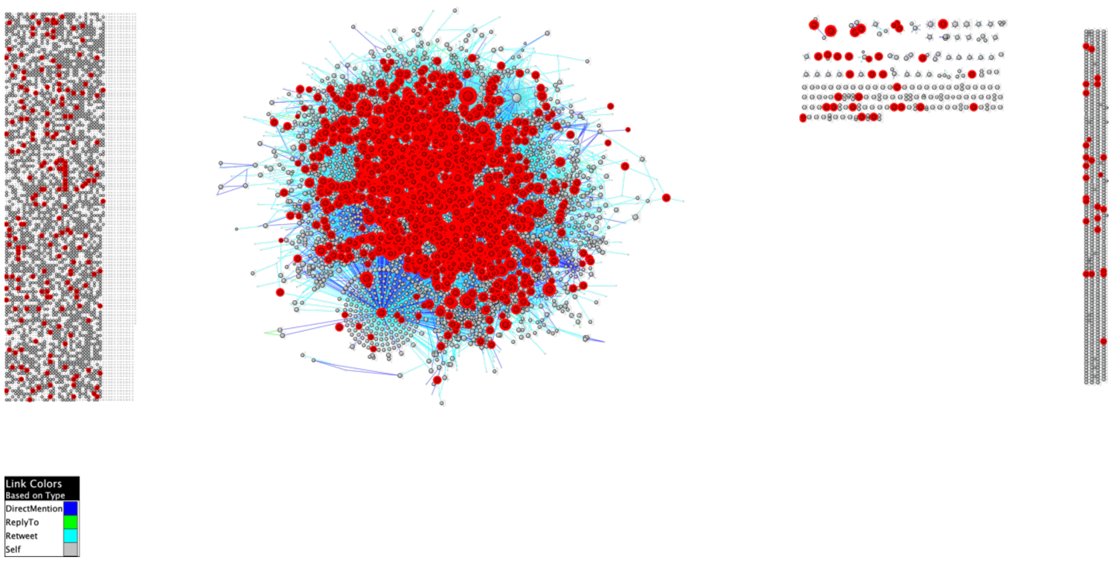


Figure 12: Network of Baltimore justice actors in T2. This network includes over 25,000 actors, with most in the main component. Actors who remain engaged in T3 are both colored red and highlighted, to improve their visibility. Nearly 90% of these engaged actors are in the main component. The main component is visible in the center, while smaller components of connected actors are shown to the right, and isolated actors are displayed on the left. Actors are connected to other actors if they mentioned, replied to, or retweeted them.

For Cleveland, the T2 network is similar in size, with 27,631 actors. A slightly higher percentage are in the main component at 78%, while, similar to Baltimore, the remaining components are small, and just 12% of actors were isolates. For this network, fewer than 4000 actors were newly engaged in justice discourse, and of those who stay engaged, an even larger percentage than observed in Baltimore—94%—were in the main component, while just 4% were isolates in T2.

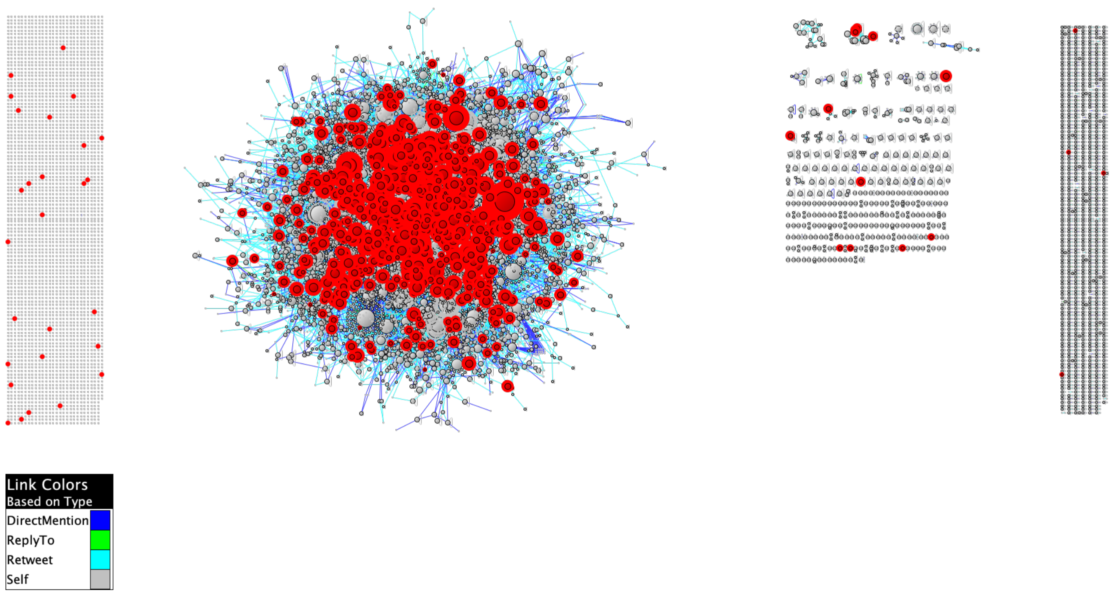


Figure 13: Network of Cleveland justice actors at T2. This network includes over 27k actors, most of whom are in the main component, as was observed in Baltimore. Actors who remain engaged in T3 are both colored red and highlighted, to improve their visibility. The main component is visible in the center, while smaller components of connected actors are shown to the right, and isolated actors are displayed on the left.

Examination of the social embeddedness dimension suggests that multiple aspects of actor network position and in- and out-degree are significant in predicting continued engagement. Several OutDegree metrics, which reflected choices by an actor to engage with or amplify other actors in the network were positive and significant. In Baltimore and Cleveland, this was observed for OutTieCount_DirectMention, while in Cleveland, OutTieCount_Retweet and OutTieCount_ReplyTo were also both positive and significant. In Cleveland, increases in Closeness and Betweenness were significant and positive. However, in Baltimore, increases in Betweenness were significant and negative.

4.6 Topical and General

As previously discussed, a number of Empath variables relating to the topics of crime, violence, the law, and conflict were considered. Only a few were significant. In Baltimore, this included references to violence, weapons, and killing. The Empath weapon variable was significant but negative in Baltimore. The achievement variable was significant and negative in Cleveland alone.

For the sake of completeness, several standard metrics of engagement with the Twitter platform, such as number of friends, followers, and statuses (tweets), were included in the model. Having more statuses had a small positive effect for both models; in other words, people who tweeted more frequently about any topic during T2 were more likely to remain engaged in justice discussions during T3. With respect to specific justice tweeting activity in T2, for Baltimore, an increase in count was positive and significant, while the opposite occurred in Cleveland.

4.7 Review of engagement models at different thresholds

Previous discussion of the models for predicting ongoing justice engagement focused on a single threshold of activity, four or more justice tweets in T3. Models were also generated at higher and lower thresholds, i.e., three or more tweets and five or more tweets. Reviewing whether variables were consistently found to be significant or not (at $p < .05$), finds substantial agreement. For Cleveland, models at all three thresholds agree 87% of the time (40 variables are never significant at these thresholds, and 11 are significant at all three thresholds). There are seven variables which are significant at one or two thresholds. These include several image variables and social embeddedness variables. Several of these were close to significance, with $p < 0.06$-0.09) at some thresholds. For Baltimore, models at all three thresholds agree 73% of the time (33 variables are never significant at these thresholds, and 10 are significant at all three thresholds). The remaining variables which are significant at one or two thresholds include several emotion variables, image variables, and social embeddedness variables. Of the ten variables that are significant at only one threshold, seven are somewhat close to significance (0.052 to 0.091) at a second threshold.

Tables for agreement of engagement models across all three thresholds for Baltimore and Cleveland are found in Appendix D, along with the complete models across thresholds.

4.8 Summary

Analysis of over 14 million tweets from Baltimore and Cleveland found that a substantial proportion of users in both cities posted about justice in the context of galvanizing events in which police actions took the lives of young African Americans in their community. While justice activity peaked during these events on Twitter, it remained above baseline levels for a significant period of time following these events. Networks of actors posting on this topic grew during T2, with most new contributors joining a large connected component linking actors together, rather than forming small components or remaining isolated. A substantial percentage of these actors remained active and connected after the initial furor of these had died down.

This research examined whether dimensions previously associated with ongoing involvement in civic and political activities, including identity, effort, emotion, and social embeddedness, also pertained in the sphere of digital engagement. Logistic regression models using variables reflecting each of these dimensions were able to predict continued engagement in justice discourse on social media in Baltimore and Cleveland. With the exception of identity, evidence supporting the remaining dimensions' utility is found, yet the results are more nuanced than might be expected. These issues, and possible explanations and interpretations, are discussed further in the next chapter.

Chapter 5: Discussion

Social media platforms like Twitter provide opportunities to engage in widespread discussions around the topic of justice. While one of the benefits of these platforms is they reflect a global community of users, they also provide opportunities for local communities to engage in reflection and sensemaking relating to their circumstance (Glasgow et al., 2014; Vieweg et al., 2010), and to push for change. This research demonstrated that it is possible to predict newcomers' ongoing digital engagement in justice discourse, using data from Baltimore and Cleveland after two galvanizing events in which police actions led to the deaths of African American community members. Building on a perspective informed by prior work on activism and engagement in social movements in the physical realm, this work focuses on the dimensions of identity, effort, emotion, and social embeddedness, as reflected in Twitter behaviors from users in these respective communities. This chapter discusses and interprets the findings for each dimension in more detail, as well as strengths and limitations to this approach in general. Before doing that, however, I first provide more background on the similarities and differences between Baltimore and Cleveland—both as American cities and in how they responded to the shootings—to begin unpacking why differences were observed across the two models.

5.1 Historical and event-driven perspective on Cleveland and Baltimore

From a historical perspective, there are a number of demographic similarities between Cleveland and Baltimore. Both cities were founded or incorporated in the late 1700s, and at their peaks grew to be among the largest cities in the United States. Both experienced an influx of African American residents during the Great Migration of the 20th century. Each city has faced economic downturns from deindustrialization and depopulation, and each experienced protests and riots during the Civil rights era. Both cities deal with the legacies of segregation, lead contamination, lack of investment and blight. Their populations have shrunk by at least one third since their peak. Both cities have substantial black populations (nearly 50% in Cleveland, and over 60% in Baltimore), and high poverty rates (approximately 30% in Cleveland and over 20% in Baltimore).[12]

Policing in both cities has a problematic and troubled history. The Department of Justice (DOJ) investigated the Cleveland Police Department (CPD), and entered into a consent decree with the city first in 2002 for issues including excessive force and racial profiling. Over a decade later, when Tamir Rice was killed, the DOJ was again investigating the CPD for a pattern and practice of excessive force, including:

- The unnecessary and excessive use of deadly force, including shootings and head strikes with impact weapons;
- The unnecessary, excessive or retaliatory use of less lethal force including tasers, chemical spray and fists;
- Excessive force against persons who are mentally ill or in crisis, including in cases where the officers were called exclusively for a welfare check; and
- The employment of poor and dangerous tactics that place officers in situations where avoidable force becomes inevitable and places officers and civilians at unnecessary risk.

The DOJ report was released on December 4, 2014, just weeks after Tamir Rice's killing, and Cleveland entered into a second consent decree.

Baltimore City invited the Department of Justice to investigate the Baltimore City Police Department after Freddie Gray's death. In 2017, the City of Baltimore and the DOJ entered into a consent decree, based on DOJ's findings that the Baltimore City Police Department (BPD) had engaged in a pattern and practice of conduct that violated the First, Fourth, and Fourteenth Amendments to the Constitution, specifically including:

[12] https://www.census.gov/quickfacts/fact/table/baltimorecitymaryland,clevelandcityohio/PST045219

- making unconstitutional stops, searches, and arrests;
- using enforcement strategies that produce severe and unjustified disparities in the rates of stops, searches and arrests of African Americans;
- using excessive force; and
- retaliating against people engaging in constitutionally-protected expression.

While both events emerged from similar contexts, there are a number of differences between them that may have shaped public perceptions and digital engagement with justice on Twitter. In Cleveland, the cause of death for 12-year-old Tamir Rice was unambiguous. He was shot twice in the abdomen by police, and died from these wounds. The shooting was captured on video. An argument for the justifiability of the shooting was raised because Tamir Rice had a toy gun in his possession, which resembled a real weapon. Officers also claimed to have perceived Mr. Rice as adult, not a child, due to his size, and to have acted in fear for their safety. In this context, a demonstration supporting police was organized in Cleveland in late December 2014, amplified by #seaofblue on Twitter and Facebook, and attended by thousands. The following figure provides an example tweet from @CLEPolice, one of several about the demonstration.

Figure 14: Tweet from the Cleveland police department during a pro-police rally in the city.

Tamir Rice's killing took place shortly before the grand jury in St. Louis County declined to indict the officer Darren Wilson who shot Michael Brown in Ferguson, MO, a controversial and widely publicized decision that ignited protest nationally. The protests and rallies held in Cleveland after the death of Tamir Rice and into December were smaller in scale than those in Baltimore, and unlike Baltimore, did not involve substantial violence, looting, large numbers of arrests, or the declaration of a curfew.

The precise circumstances, cause of death, and the role of police officers in the death of Freddie Gray in Baltimore were murkier. Freddie Gray, a young Black man, was taken into custody after being arrested on charges of questionable legality.[13] He was transported, unbelted and later shackled, in a police van for roughly 45 minutes before being removed from the van, unresponsive and seriously injured, and transported to a hospital. Debate swirled about where and how he sustained his injuries—during his arrest, as a consequence of deliberately hurling himself around the back of the van, as a tragic accident from failure to properly seatbelt him, or as the result of a deliberate "rough ride" in which officers drove erratically to cause injury to a suspect. Freddie Gray died a week later from his injuries. An autopsy found Freddie Gray suffered a single "high-energy injury"—similar to a shallow-water diving accident—most likely caused by the police van rapidly decelerating, and declared his death a homicide.

Large—and largely peaceful—protests in Baltimore followed Freddie Gray's death. However, after his funeral, this changed. Police, responding to social media rumors of a "purge" starting at Mondawmin Mall and travelling downtown, confronted students after school. The situation rapidly devolved into violence, rioting, and the deployment of teargas in the

[13] Freddie Gray was charged with possession of an illegal switchblade. Further examination of the knife found it was not a switchblade, and was legal. It is possible that the arresting officers were simply mistaken about the knife. The issue was not pursued at trial.

neighborhood.[14] This was followed with further destruction, arson, and looting that damaged dozens of buildings and vehicles in the city. Hundreds of people were arrested, and officers were injured in the conflict. Media attention to the rioting and destruction seemed to eclipse attention on the circumstance of Freddie Gray's death and the longstanding issues of the operation of the criminal justice system in Baltimore.

Unlike Cleveland, authorities in Baltimore moved swiftly to take action against the officers involved in Freddie Gray's arrest and transport. By May 1, 2015, the state's attorney, Marilyn Mosby, announced multiple charges against six officers, which may have helped dampen ongoing unrest. A grand jury returned indictments by the end of the month. However, trials of the first three officers ended in a hung jury and acquittals, after which charges against the remaining officers were dropped. Many supporters of the officers found vindication in this outcome, while others perceived it as further evidence of a deeply flawed criminal justice system in the city. No pro-police rallies or demonstrations of the scale seen in Cleveland were organized in Baltimore during the period after Freddie Gray's death.

To summarize, both Baltimore and Cleveland experienced early periods of growth and relative prosperity, but later saw a dramatic downturn in fortunes. Each city now struggles with a shrinking population and tax base, as well as legacies of segregation, environmental degradation, lack of investment and blight. Policing and the functioning of the criminal justice system in each city have been deeply flawed and problematic, featuring systemic abuses that have led to Department of Justice consent decrees. Protests and civic engagement in both cities after the killings of the two young men ignited national and international attention. In Baltimore,

[14] https://www.baltimoresun.com/news/crime/bs-md-ci-freddie-gray-violence-chronology-20150427-story.html

but not Cleveland, the situation escalated to violence and rioting. Efforts to hold the cities, their police officers, and their police departments accountable were met with mixed results. Families of the victims in both cases received substantial civil payments. Rapid efforts to employ the criminal justice system against the officers in Baltimore faltered within months, while less-aggressive efforts in Cleveland dragged on for years. While the need for reform in both cities seems evident and well-established, there are portions of the populations in each city who supported the established operations of the criminal justice system, did not perceive the galvanizing events as signals of larger problems, and organized and expressed support for police on social media. This constellation of historical and contemporary factors, as well as two galvanizing events in which African Americans were killed through police actions, against a backdrop of growing national awareness of injustice and systemic bias in policing, led to environments in both cities where increased and continuing engagement in justice discourse on social media could be expected. However, differences in the details of the events and official response could shape the response.

5.2 The role of identity variables in predicting continued justice engagement

Modifications to elements of a user profile can be used to signal changes in identity from the personal level to affiliation and alignment with broader audiences (Schwämmlein & Wodzicki, 2012), to participation in social movements (Raynauld et al., 2018). For example, a Twitter user may add #blacklivesmatter to the end of the username or their profile to signal solidarity with the wider movement. Recent work has quantified frequency of updates to Twitter profiles (Wesslen et al., 2018), with 2% (screen name) to over 15% (profile summary) of active Twitter users updating an aspect of their profile over the course of eight weeks as of late 2017. In Baltimore and Cleveland, it was not a common practice for the users we are studying to make identity-related changes to their profiles. For Baltimore, a few users were observed to change their user names *to Freddie Gray 🎗, #BlackLivesMatter*, and *JusticeForFreddie*. Similarly in Cleveland, *No Justice No Peace, Tamir Rice,* and *#BlackLivesMatter* also appeared. These profile changes were relatively infrequent and not predictive of ongoing engagement with justice. During the timeframe of the data collection, relatively few users updated their profiles to highlight or signal justice issues.

There are many possible reasons for the limited engagement in these types of identity changes (RQ1). Individuals may perceive constraints on updating their public profile identity to reflect engagement with justice. For example, a user's account may reflect a professional as well as a personal identity, raising potential challenges to self-branding, self-representation, and employer reputation (Hanusch & Bruns, 2017). Changes to user display names and screen names, profile descriptions and images, may be perceived as more effortful and more permanent than posting a tweet on the topic; tweets quickly fade into the larger stream of tweets a user posts, whereas profile information remains static and consistent. Twitter has fewer affordances and more friction for displaying aspects of identity compared to a platform like Facebook, where the click of a button can update a profile image to include a photo frame showing support for a social cause (Wilson & Cohen, 2019). Because of this, Twitter users may be less likely to engage in the changes to identity presentation considered in RQ1. Further, the complex process of learning about an unfolding, galvanizing event, sharing information about actions of others (e.g., police or other agents of the criminal justice system), as well as reflections and comments situating one's self in this context, may muddy the informativeness of pronoun usage as an indicator of evolving digital identity.

5.3 The role of effort variables in predicting continued justice engagement

RQ3 considers whether producing (and sharing) more effortful posts—those containing hashtags or images—is predictive of ongoing engagement in justice discourse. The presence of images or hashtags in tweets has been shown to increase audience engagement on Twitter.[15] In the context of disasters, such as 2012's Hurricane Sandy, users incorporated images in posts for reasons beyond information sharing, reflecting their vantage points as disaster victims rather than those of official responders, suggesting "possible changes in the politics of representing disasters—a potential turn from top-down understandings of disasters to bottom-up, citizen informed views" that reshape the social experience of disasters (Murthy et al., 2016, p.113). Similarly, hashtags in a geographically focused movement, Occupy Oakland, were observed to play a number of roles, including supporting live-tweeting to document the event from a participant rather than official perspective, enabling solidarity actions, and providing a crucial connection to those not physically present both before and after police dispersion of activists (Croeser & Highfield, 2014).

In both cities, sharing posts with images increased the probability of continued engagement in T3. The effect of using hashtags relating to justice or the victims proved more complex. While justice hashtags had positive predictive value in Baltimore, this was not seen in Cleveland. Some justice hashtags observed in the Cleveland dataset semantically appear to reflect a position calling for reform of the criminal justice system and actions of the police (e.g., #blacklivesmatter, #icantbreathe, #nojusticenopeace), while others seem to align with support of the status quo with respect to the criminal justice system, centering of police perceptions and perceived threats to officer safety, and framing of protest as anti-law enforcement. These include

[15] https://blog.twitter.com/official/en_us/a/2014/what-fuels-a-tweets-engagement.html

#alllivesamatter, #policelivesmatter, #seaofblue, and #thinblueline. However, the usage of these sets of hashtags is more nuanced and complex. As observed by Gallagher et al., contending hashtags have contentious dynamics (Gallagher et al., 2018). Hashtag hijacking or content injection occurred in justice discourse, as well as deeper content commentary and content engagement. This was perhaps amplified in Cleveland by the planned #SeaofBlue rally in support of police. The following tweets are examples of this phenomenon.

- *@A, I'm saying instead of #BlackLivesMatter #WhiteLivesMatter #MuslimLivesMatter #ChristianLivesMatter*
- *17 police officers have been shot now in line of duty since Michael Brown.... #BlueLivesMatter #TheyCantBreathe*
- *RT @B: If you think #AllLivesMatter then stop ignoring the oppression & brutality faced by Black communities. #BlackLivesMatter*
- *RT @C: I don't understand why people don't see that #alllivesmatter is just a nicer way of saying #shutupaboutracism.*
- *RT @D, @E @F #bluelivesmatter rally in CLE will be a demonstration of white privilege #hijackedmessage*

It may be the case that participants in justice discourse who also perceive the issues and events as evidence of a need for deeper dialog with communities and institutional change are more likely to remain engaged over time. Those participants who are satisfied with and supportive of police actions in general may be less likely to sustain engagement after the peak of galvanizing events that challenged the legitimacy and fairness of law enforcement and appeared to demonstrate failures of procedural justice. Future research should work to unpack these complex relationships.

5.4 The role of emotion variables in predicting continued justice engagement

While anger has been called the prototypical protest emotion (van Stekelenburg, 2013), the role of emotions in activism has been characterized as multifarious, and shifting, contributing to both the functioning and sustainability of activism (Brown & Pickerill, 2009). These emotions relevant to activism have been described as reflecting 'transitory responses' to external events (anger, indignation, fear), as well as 'underlying affects' (loyalty to family, friends or nation; or fear of others) (Jasper, 1998).

Other positive emotions can be evoked in protests that incorporate theatrical or entertaining elements (Brown & Pickerill, 2009), which are then shared via social media. Exploring RQ3, we found evidence of this in the Baltimore dataset, where protests at the killing of Freddie Gray incorporated music and dance, at time turning the streets into a block party.

- *RT @A: PEACEFUL MICHAEL JACKSON DANCE PARTY HAS ERUPTED IN #BALTIMORE. #FreddieGray http:// ..*
- *This impromptu dance party in #Baltimore might not make it to most news outlets. #FreddieGray #BaltimoreUprising http://..*
- *RT @lukebroadwater: .@baltimoresun's Kevin Richardson shoots the celebratory dance party going down at City Hall #FreddieGray http://..*
- *It really feels more like a block party than a protest here at North and Penn where the crowd from the #FreddieGray rally ended up*

However, positive emotions in general were not predictive of ongoing engagement.

Social media platforms may also amplify expressions of hostility, racism, and irony and sarcasm, while expressions of compassion or empathy may not experience the same treatment (Nikunen, 2018). The data examined in this research covers a broad spectrum of activity, encompassing more than protest. In the context of the galvanizing and contentious events in Baltimore and Cleveland, interpretation of the emotion dimension is challenging.

Both events evoked intense emotion for some individuals. Responses of grief, anger, sadness, and fear appeared, as did love, support, and hope. Multiple emotions can be expressed in a single tweet. The complexity of human emotional expression in short natural language texts remains a challenge. Sarcasm is seen in some cases ostensibly expressing positive emotion, as is the idiomatic usage of “trigger happy.” The following are a few illustrative examples of justice tweets from Cleveland that contain positive emotion, based on Empath.

- *Glad to see on the news that Cleveland cops are now wearing body cameras. Maybe we can start avoiding unnecessary police brutality.*
- *Gotta love the @clevelanddotcom comment septic tank. Loads of white racist retired suburbanites weighing in on #CLE police policy as if..*
- *All these anti police people who are glad this happened are all burdens to society and will be living off the govt forever. #Coincidence??*
- *RT @A: Police are so trigger happy they're shooting 12 year old kids and you want us to believe you're the ones to protect us?*
- *RT @B: Lesley McSpadden comforted by family as public learns her son's killer will not be charged (A Photo) http://...*

Considering RQ2, we see that while expressions of emotion in post content could be predictive of ongoing digital engagement in Baltimore, there was no simple pattern to observation. Expression of negative emotion was negatively associated with continued engagement, while expression of horror was positively predictive. The complex role of emotion in digital engagement in the context of contentious events, the limitations of Empath, and potentially the algorithmic amplification by the platform of some types of emotive content over others, may all contribute to this observation.

<u>5.5 The role of topical variables in predicting continued justice engagement</u>

Several topical variables, as computed by Empath, were among the significant predictors of sustained engagement, particularly in Baltimore. These variables were related to crime and violence. Differences in significance may reflect details of the galvanizing event, or police response to it.

For Baltimore, “weapon” was positively predictive of engagement. The justification for Freddie Gray’s arrest was that he was found in possession of an illegal weapon. The rioting and looting in Baltimore also at times involved weapons (rocks, bricks, and riot control devices), and may have also made "stealing" more salient in justice discourse. Each of the example tweets below was positive for Empath "weapon" and was retweeted over a dozen times to nearly one hundred times.

- *RT @A: Fire beside Pratt library was not caused by Molotov cocktail. The teargas grenade landed on trash and its sparks set the fire. Watched it.*
- *RT @B: Marilyn Mosby says knife officers found on Gray was not a switchblade & was legal, making his arrest illegal #FreddieGray*
- *RT @C: What is throwing bricks and rocks at police going to solve. Not a got damn thing.*
- *RT @justin_fenton: Police shot down the sidewalk with what I believe to be pepper bullets, as pepper filled the air. Stings the eyes, throat*

In contrast, in Cleveland, fewer justice tweets mentioned weapons. While references to the toy gun carried by Tamir Rice and to the police shootings did appear, none of them were retweeted more than a few times. The figure below provides an example.

Figure 15: Justice tweet from Cleveland that referenced the Empath concept "weapon"

5.6 The role of social embeddedness in predicting continued justice engagement

Social connectedness matters for collective and connective action, and to perceptions of justice (Cropanzano, Ambrose, Masterson, et al., 2015). Having ties and interactions in social networks which discuss political issues increases an individual's propensity to participate in politics (Mcclurg, 2003). Observations that contentious political participation is increasing while embeddedness in traditional mobilizing structures (e.g., political parties, unions, and formal organizations) is decreasing, attribute this seeming paradox to the growing importance of ties to digital communities or networks engaged around these issues (Stekelenburg et al., 2013). At an individual level, this social embeddedness provides paths to inform and educate, influence, motivate, and mobilize. At a macro level, it can lead to political outcomes, increased media coverage (of protests or issues), and shift public opinion and awareness (Stekelenburg, 2012). Social embeddedness, as measured through network ties, has also been shown to sustain ongoing digital engagement after peaks of activity tied to significant events (Leong et al., 2019).
RQ4 examines an actor's social embeddedness, as measured by social network metrics. In both Baltimore and Cleveland, a number of social network metrics used to capture social embeddedness were predictive of ongoing engagement. In both cities, an OutDegree metric reflecting how many other an actor has mentioned, OutTieCount_DirectMention, was positively predictive. Several other metrics which measure an actor's position in the overall network structure but are not exclusively dependent on the actor's behavior, as OutDegree measures are, were significant. In Baltimore, this was observed for Betweenness. While in Cleveland, Degree was positive and significant, this was not observed for Baltimore. Closeness centrality, which captures social embeddedness in terms of possessing a central position with short distances to

others in the network was positively predictive for Cleveland. With respect to RQ4, it appears social embeddedness, measured by a variety of social network metrics, was predictive of ongoing engagement in these two cases.

5.7 Social embeddedness and justice tweeting

With respect to social embeddedness, one could consider whether justice tweeting in itself accounts for an actor's social embeddedness. There are number of social network analysis metrics considered in this study, some of which reflect emergent properties of the overall network. Agency by an actor cannot control these. For example, the actor controls their outdegree and type of tie—which others they choose to reply to, mention, or retweet. Closeness, betweenness, and indegree metrics depend on the actions and behaviors of the rest of the network, and the resulting network structure that emerges. When an actor chooses to connect with others through their tweets, that behavior alone cannot ensure the actor will belong to the main component. In both the Baltimore and the Cleveland cases, there were components that contained over 20 connected actors that still fell outside the main component. One actor had over 50 justice tweets in T2, yet still was not in the main component. Thus, an actor can have substantial activity in T2 without that activity translating into a central position by many social embeddedness metrics. Engagement (based on tweets) in T2 probably sets a ceiling for outdegree-based aspects of social embeddedness, as the character limit for tweets may make difficult to mention more than 10 unique other people in a single tweet. The floor is zero, as many tweets don't mention others.

Nevertheless, there is a potential element of reward associated with being acknowledged or responded to by other members of the community. If an individual is posting tweets that garner a positive response and praise from other (especially high-status) members of the community, that may encourage continued activity. If, instead, an actor receives criticism or condemnation for their posts, that might discourage further activity, unless the actor is a troll who is trying to offend people and garner negative response.

Prior work suggests actors in a social network possess a relatively limited and local view of the overall network structure, with a "horizon of visibility" of one to perhaps two hops out (Friedkin, 1983). Review of the justice content in both Baltimore and Cleveland reveals tweets and actors with very different framings of the events (e.g., support to victims of police violence and importance of reform vs. support to police and maintenance of the status quo). Actors interacting with like-minded others may not be aware of activity outside of their local neighborhood, or have a sense of their overall position in the network.

5.8 *Limitations*

Data for this research relied on the Twitter Powertrack. While this is a powerful method for collecting large amounts of data from a city, it could miss some residents, particularly if there was no profile information or enrichment to associate them with the city during the timeframe. The affordances of the Twitter platform for expressing each of the dimensions, particularly identity, are limited. Character limits on the length of tweets leads to short texts which may pose challenges to text processing. For example, tweets are analyzed independently while users may string together a series of tweets (thread) on a topic.

While the use of machine learning to identify relevant justice tweets does improve retrieval over popular methods like hashtag analysis, it is not infallible. In particular, since a Twitter dataset can contain many copies of a popular tweet (e.g., multiple retweets of the same initial tweet), a classification error, such as a false positive, will be repeated on every instance of that tweet.

Another limitation is the length of the baseline for determining users who were already engaged in justice discourse prior to the galvanizing event. It is possible that some prior participants in this discourse were not active during the baseline.

While this research examined millions of tweets from two cities that experienced killings of Black men, this type of event, tragically, has happened hundreds of times in the United States, in hundreds of communities. Insofar as unique aspects of a city's history, the event itself, and the response of authorities are likely to shape the justice discourse that happens, findings from these two cities may not fully generalize to other incidents. Future research should continue to explore factors that may hold consistent across events vs. factors that vary based on qualities of the city and qualities of the event.

Further, digital engagement for members of a community can occur across multiple social media platforms. This research focused solely on Twitter-based engagement. Social media posts, for all their potential power, reach, and ability to shape narratives and drive action, are still often pale reflection of the richness, multifacetedness, or depth of user's underlying beliefs and values relating to complex issues such as justice. Future researchers should consider ways to study these questions through multiple methods, perhaps using social network data in conjunction with qualitative methods. One example of this to highlight is work by Deen Freelon and colleagues on the early use of Twitter as part of the Black Lives Matter movement (Freelon et al., 2016)

Chapter 6: Conclusion

Examination of over 14 million tweets from Baltimore and Cleveland found that a substantial proportion of users in both cities posted about justice in the context of galvanizing events in which police actions took the lives of young African Americans in their community. These contentious events sparked sadness, outrage, and demands for change to the justice system, as well as outpourings of support for police actions. Justice activity peaked during these events on Twitter and remained elevated afterward. Networks of actors posting on this topic grew during the peak, with most new contributors joining a large connected component linking actors together. A substantial percentage of these actors remained active and connected after the initial furor of these events had died down.

This research considered a set of factors to predict Twitter users' ongoing engagement in justice discourse following a galvanizing event, exploring the role of identity, emotion, social embeddedness, and effort. Logistic regression models using variables plausibly reflecting each of these factors were able to predict continued engagement in justice discourse on social media in Baltimore and Cleveland. With the exception of identity, evidence supporting the remaining dimensions' utility is found, yet the results are more nuanced than might be expected, with some differences observed in the specific significant predictive variables within each dimension, and the magnitude and direction of each. In some cases, distinct aspects of each incident (e.g., manner of death of the victim, presence or absence of riots, actions by authorities against the officers) appeared to affect the results.

Using machine learning to determine which tweets concerned justice is an advance over more typical hashtag or term-based methods. This issue was examined in prior work on Baltimore Twitter data for this time period, which found that a hashtag-based approach for identifying relevant justice content identified a fraction of that content, underrepresented baseline activity, and failed to capture posts by some prominent actors in the network (Glasgow et al., 2020). While hashtags can be valuable indicators, a hashtag-focused approach risks systematically distorting our understanding of digital engagement with contentious social issues such as justice. A substantial number of participants in justice discourse in both Cleveland and Baltimore did not employ justice hashtags or narrow key terms in their relevant tweets.

Incorporating images in this classification more fully captured the meaning of each tweet, and they may have utility for analysis of primarily visual social media platforms. This type of approach could benefit investigation of topics which cannot easily be reduced to a small number of terms of hashtags, or those in which a hashtag has multiple meaning or referents (e.g., #BLM can refer to Black Lives Matter or the Bureau of Land Management, #HRC can refer to Hillary Rodham Clinton or the Human Rights Council). As a hashtag may inherently frame an issue or take a stance on it, this approach also helps mitigate potential biases introduced by selecting data from that frame, if that is not appropriate for the research question. Similarly, users, whether individuals or organizations, may avoid such charged language in their posts, by personal choice or organizational policy.

6.1 Contributions of this study

This dissertation makes several contributions and extensions to literature on social movements and social media, and it offers methods considerations for future event-based studies of social media data. It broadly demonstrates the applicability of a computational social science approach (Lazer et al., 2009) to understanding digital engagement with justice. This approach extends more traditional data gathering on justice-related topics from surveys and interviews (Engel, 2005; Hurwitz & Peffley, 2005) or experiments on small sets of actors (Lind et al., 1998), by using data derived from social media behaviors of participants from the community during an event. In this section, I highlight how this study provides new insights into factors that lead to increased participation in social justice discourse in communities directly affected by police violence; the challenges of using hashtags alone to capture a complex social phenomenon; the benefits of including both psychological and network-based measures when considering engagement on a platform like Twitter; and the role that weak ties may play in promoting and extending social justice discourse. More specifically, my dissertation contributes to the extant work by:

- Demonstrating that it is possible to predict ongoing engagement in justice discourse on social media
- Extending prior work assessing digital engagement and social justice by focusing on local contexts
- Providing empirical support for the limitations of hashtag-based analyses of complex political and social phenomena
- Providing empirical support that weak ties in online networks contribute to ongoing civic engagement

The remainder of this section provides more detail and context to these contributions.

6.1.1 Predicting ongoing engagement in justice discourse on social media

Our understanding of the nature of activism and civic and political engagement is being reshaped in the digital age (Freelon, 2014; Zuckerman, 2014). As part of this research, I developed an approach to enable prediction of ongoing social media engagement around justice, then demonstrated its prediction utility. This approach is informed by work in the social sciences that explains participation in social movements, activism, civic engagement, and related phenomena (Klandermans, 1996). While survey and interview methods have been used to gather data with respect to these phenomena in physical spaces, I adapted the dimensions of identity, emotion, effort, and social embeddedness to digital engagement via social media posts. My analyses demonstrate an approach to (1) recognizing indicators of identity and identity change in social media, (2) analyzing social media text and emoji for indicators of emotion, (3) distinguishing aspects of posts that suggest more effort (cognitive or physical), and (4) constructing social networks that emerge from the aggregation of justice discourse activity across the community of participants. Taking this approach allowed me to both predict ongoing individual-level engagement and to assess whether each dimension contributes to predicting ongoing engagement in social justice discourse for these cases.

Additionally, I demonstrated the generalizability of these findings from one high-salience, galvanizing local event—the death of Freddie Gray in Baltimore—to the shooting of Tamir Rice by police in Cleveland. Informed by observations on the Freddie Gray case, the second case followed a similar approach, and was similarly able to predict ongoing engagement with justice discourse after a galvanizing event. By analyzing more than one case, this study highlights that while some factors remain predictive in different contexts, differences in the specifics of the galvanizing events will shape local responses, including who remains engaged over time. The killing of Freddie Gray in Baltimore, which happened after the killing of Tamir Rice in Cleveland, evoked a stronger online response, as well as a fundamentally different physical response—with riots, looting and mass arrests by police (Chapter 4 and Section 5.1). This is consistent with literature that suggests that, over time, reactions in the digital environment are shaped by accumulated frustrations over prior incidents of injustice, and are also playing out in a digital context informed by prior social media campaigns (Bonilla & Rosa, 2015).

6.1.2 Focus on localized response to injustice

Prior work examining issues of injustice, social movements, and civic action as expressed on Twitter has assessed response to events in a global context (Conover et al., 2013; Freelon et al., 2016; Ray et al., 2017; Raynauld et al., 2018). This dissertation extends prior work by focusing on the context where the injustice occurred. Specifically, it analyzes a large corpus of Twitter activity and events at the local (city) level, as experienced and reflected by the actual community going through a devastating event that, for many, crystallized ongoing issues of justice and policing within their respective cities. By providing a city-centric view on the social media response to these high-profile events, this dissertation contributes to our understanding of how localized responses unfold within the community in the immediate aftermath of galvanizing events. Prior work has shown that local response, particularly in the context of disasters and mass traumatic events, is distinct from non-local response, and better tailored to support local information needs and social cognition (Glasgow et al., 2014; Starbird & Palen, 2011; Vieweg et al., 2010).

Considering the specific events discussed in this dissertation, prior work analyzing social media discourse following the death of Freddie Gray similarly considered global patterns of activity (Welles & Jackson, 2019). That work, examining what was dubbed the "the Battle for #Baltimore," looked at Twitter activity during the timeframe of the riots in Baltimore based on two hashtags, #BaltimoreUprising and #BaltimoreRiots, analyzing hundreds of thousands of tweets to explore networked counterpublics and the social networks emerging from this discourse. The first of these hashtags was used to anchor a framing of events in Baltimore as a legitimate expression of Black protest, with unrest as emerging from how the police interacted with the community in violation of civil rights, and with mainstream media playing up violence and downplaying genuine issues of the community. The second framing, #BaltimoreRiots, emphasized the unrest in Baltimore as dangerous and unreasonable, featuring violence being perpetrated by criminals.

Taking a city-centric view of Baltimore in my dissertation helps extends our understanding and provides additional context to Jackson and Welles' (2019) findings. For example, looking at my dataset, it quickly becomes clear that the vast majority of the "battle for Baltimore" content did not arise from Baltimore itself, from the individuals and communities actually experiencing and reporting on the unrest and violence (Section 3.5.1). That said, the two framings described by Jackson and Welles (2019) were prominent in the Baltimore justice data, though their expression was not constrained to tweets containing those hashtags (Section 5.3). Additionally, their observation that references to the riots occur with counterpublic interpretations that are consistent with the uprising framing vs. the violent criminal framing was also supported by the Baltimore justice data. Therefore, by combining the findings from this dissertation with prior work on the same event, we can develop a more holistic understanding of how social media was used to discuss the event and identify contextual factors that might influence outcomes.

6.1.3 Limitations of only using hashtags in building social media datasets

Prior work has raised concerns about potential limitations to hashtags in the study of complex or contentious social and political phenomena (Mahrt & Scharkow, 2013; Olteanu et al., 2015; Tufekci, 2014), including issues of representativeness, completeness, and reductiveness. While this study was not focused on evaluating hashtag usage in justice discourse specifically, it does contribute insights in this area. It demonstrated that, in these two galvanizing events, hashtags coined to capture a demand for justice for the victims of police violence (#justiceforfreddiegray and #justicefortamirrice) appeared relatively infrequently, no more than a few thousand times. Even #blacklivesmatter, the hashtag identified with the movement associated with criminal justice reform and an end to police violence against African Americans, occurred fewer than 10,000 times (Sections 3.5.1, 3.5.2). This study thus contributes additional support to the prior observations that the "hashtag we focus on does not cover all the discussions and contributions around the issue at core" (Olteanu et al., 2015) and suggests a need for more robust methods for identifying relevant social media content.

This study used machine learning to detect justice discourse broadly, rather than relying on a few relevant hashtags. Numerous features capturing tweet images, emoji, and hashtags, as well as text were considered in an effort to improve recall and avoid potential biases that could introduced by differential hashtag adoption across the community, something Tufekci (2014) raised in her analysis of hashtags pertinent to political 2011 uprisings in the Middle East. She observed that use of prominent hashtags may reflect political preferences of supporters or opponents to a cause. If hashtag users are thus prone to selection biases, this complicates analysis. My dissertation contributes two cases pertaining to criminal justice and police violence in the United States; findings support the idea that recall about an event or issue can be improved—and the potential effect of specific relevant hashtags being embedded in particular cultural or political frameworks can be mitigated—by an approach to data collection and determining relevance that is not exclusively reliant on one or a few hashtags.

6.1.4 Clarifying the role of weak ties in justice engagement

Prior work using survey data has shown that "weak ties" in online networks have a strong relationship with civic involvement (de Zuniga et al., 2010). This study, by capturing justice networks and changes to justice networks over time in social media, contributes to our understanding that “weak ties” play an important role with respect to ongoing engagement in justice discourse (Section 4.5, Section 5.6). Notably, actor position in justice discourse networks and patterns of interaction can be predictive of future engagement in justice discourse. This provides additional support and extends prior work looking at the network structures of a youth organizing group working on social justice issues (Evans, 2013), and provides evidence that providing structural knowledge of social media networks has the potential to identify future outreach targets. The social embeddedness factor explored in this dissertation, the social network metrics supporting it, and their connection to ongoing engagement could suggest additional avenues for outreach.

6.2 Future research and recommendations

This research has demonstrated that it is possible to adapt factors associated with ongoing engagement in civic activity and social movements in physical space to predict ongoing digital engagement on Twitter. Extension of this approach to justice discourse on other platforms, particularly platforms with a broader or richer set of mechanisms to support expression of identity or emotion, could further clarify the importance of these factors, assuming adequate datasets could be collected from those platforms. The exploration of alternative text processing or classification methods might contribute to improved accuracy. Determining if engagement in justice discourse continues over extended periods of time, such as a year after an event, would further inform our understanding of how enduring these changes are.

Replication of this approach on other events, might clarify patterns of predictive variables within dimensions, and illuminate if particular aspects of an event shape which variables are predictive or influence overall rates of ongoing engagement. For example, the killing of George Floyd by Minneapolis police in 2020 has numerous parallels with the two galvanizing events studied in this research. It occurred in a city with similar historical and demographic backgrounds, and facing similar challenges to Baltimore and Cleveland. The event in Minneapolis led to massive, sustained, nationwide protests during a pandemic, which reasonably would have been expected to suppress protest activity. The officer Derek Chauvin has been charged and is, at the time of this writing, standing trial for the killing of George Floyd. In contrast to Baltimore and Cleveland, there is both abundant video evidence documenting the killing, and numerous eyewitnesses to it who are testifying at the trial.

The Twitter platform offers relatively few mechanisms for expression of identity, particularly in static (profile) form. However, alternative designs that enabled multiple, low friction ways to express aspects of identity or affiliation with causes or movements might better support civic engagement through social media platforms. Similarly, a broader repertoire of reaction options (beyond a “like” or “heart”) could enable more contextually sensitive and appropriate responses to posts relating to traumatic, painful, or contentious events, and also provide a clearer window into the user’s emotional reaction.

Pursuing the research directions proposed above would support a robust path forward for understanding the roles of the factors of identity, emotion, effort, and social embeddedness in digital engagement around contentious issues like justice.

Appendix A: Emoji classes

Emoji	Person	Emotion
emoji_face_with_tears_of_joy	Yes	Positive
emoji_smiling_face_with_smiling_eyes	Yes	Positive
emoji_loudly_crying_face	Yes	Negative
emoji_smiling_face_with_heart_shaped_eyes	Yes	Positive
emoji_weary_face	Yes	Negative
emoji_skull		Negative
emoji_unamused_face	Yes	Negative
emoji_face_throwing_a_kiss	Yes	Positive
emoji_person_raising_both_hands_in_celebration	Yes	Positive
emoji_heavy_black_heart	Yes	Positive
emoji_white_smiling_face	Yes	Positive
emoji_smiling_face_with_sunglasses	Yes	Positive
emoji_eyes	Yes	
emoji_clapping_hands_sign	Yes	Positive
emoji_grinning_face_with_smiling_eyes	Yes	Positive
emoji_smirking_face	Yes	Positive
emoji_person_with_folded_hands	Yes	
emoji_relieved_face	Yes	Positive
emoji_smiling_face_with_open_mouth_and_cold_sweat	Yes	Positive
emoji_heavy_exclamation_mark_symbol		Positive
emoji_ok_hand_sign	Yes	Positive
emoji_sleeping_face	Yes	Ambiguous
emoji_flushed_face	Yes	Negative
emoji_information_desk_person	Yes	
emoji_neutral_face	Yes	
emoji_confused_face	Yes	Negative
emoji_party_popper		Positive
emoji_pensive_face	Yes	Negative
emoji_expressionless_face	Yes	
emoji_two_hearts	Yes	Positive
emoji_face_savouring_delicious_food	Yes	Positive
emoji_tired_face	Yes	Negative
emoji_open_hands_sign	Yes	Positive
emoji_flexed_biceps	Yes	Ambiguous
emoji_smiling_face_with_horns	Yes	Positive
emoji_face_with_look_of_triumph	Yes	Positive
emoji_face_with_stuck_out_tongue_and_tightly_closed_eyes	Yes	Positive
emoji_face_with_stuck_out_tongue_and_winking_eye	Yes	Positive
emoji_smiling_cat_face_with_heart_shaped_eyes		Positive

emoji_thumbs_up_sign	Yes	Positive
emoji_smiling_face_with_open_mouth_and_tightly_closed_eyes	Yes	Positive
emoji_sparkles		Positive
emoji_pouting_face	Yes	Negative
emoji_raised_fist	Yes	Negative
emoji_victory_hand	Yes	Positive
emoji_kiss_mark	Yes	Positive
emoji_winking_face	Yes	Positive
emoji_fisted_hand_sign	Yes	Ambiguous
emoji_blue_heart	Yes	Positive
emoji_sleepy_face	Yes	Negative
emoji_face_with_cold_sweat	Yes	Negative
emoji_crying_face	Yes	Negative
emoji_happy_person_raising_one_hand	Yes	Positive
emoji_waving_hand_sign	Yes	
emoji_face_with_stuck_out_tongue	Yes	Positive
emoji_thumbs_down_sign	Yes	Negative
emoji_dancer	Yes	
emoji_purple_heart	Yes	Positive
emoji_man_and_woman_holding_hands	Yes	Positive
emoji_face_screaming_in_fear	Yes	Negative
emoji_smiling_face_with_open_mouth_and_smiling_eyes	Yes	Positive
emoji_face_with_medical_mask	Yes	
emoji_heart_with_arrow	Yes	Positive
emoji_sparkling_heart	Yes	Positive
emoji_face_with_no_good_gesture	Yes	Negative
emoji_tongue	Yes	
emoji_black_sun_with_rays		Positive
emoji_confounded_face	Yes	Negative
emoji_white_right_pointing_backhand_index	Yes	
emoji_broken_heart	Yes	Negative
emoji_raised_hand	Yes	
emoji_revolving_hearts	Yes	Positive
emoji_disappointed_face	Yes	Negative
emoji_person_bowing_deeply	Yes	
emoji_black_heart_suit	Yes	Positive
emoji_grinning_face	Yes	Positive
emoji_smiling_face_with_halo	Yes	Positive
emoji_growing_heart	Yes	Positive
emoji_two_women_holding_hands	Yes	Positive
emoji_pile_of_poo		Negative
emoji_cat_face_with_tears_of_joy		Positive

emoji_yellow_heart	Yes	Positive
emoji_persevering_face	Yes	Negative
emoji_white_left_pointing_backhand_index	Yes	
emoji_smiling_face_with_open_mouth	Yes	Positive
emoji_grimacing_face	Yes	Negative
emoji_green_heart	Yes	Positive
emoji_runner	Yes	
emoji_woman_with_bunny_ears	Yes	Positive
emoji_confetti_ball		Positive
emoji_white_down_pointing_backhand_index	Yes	
emoji_woman	Yes	
emoji_kiss	Yes	Positive
emoji_anger_symbol		Negative
emoji_baby_angel	Yes	
emoji_kissing_face_with_closed_eyes	Yes	Positive
emoji_disappointed_but_relieved_face	Yes	Negative
emoji_bomb		Ambiguous
emoji_squared_ok		Positive
emoji_warning_sign		Negative
emoji_angry_face	Yes	Negative
emoji_baby	Yes	
emoji_face_massage	Yes	Positive
emoji_white_heart_suit	Yes	Positive
emoji_face_without_mouth	Yes	
emoji_fearful_face	Yes	Negative
emoji_princess	Yes	Ambiguous
emoji_clinking_beer_mugs		Positive
emoji_beating_heart	Yes	Positive
emoji_anguished_face	Yes	Negative
emoji_face_with_ok_gesture	Yes	Positive
emoji_police_officer	Yes	
emoji_two_men_holding_hands	Yes	Positive
emoji_white_right_pointing_index	Yes	
emoji_couple_with_heart	Yes	Positive
emoji_worried_face	Yes	Negative
emoji_white_up_pointing_index	Yes	
emoji_face_with_open_mouth_and_cold_sweat	Yes	Negative
emoji_guardsman	Yes	
emoji_man	Yes	
emoji_boy	Yes	
emoji_white_up_pointing_backhand_index	Yes	
emoji_astonished_face	Yes	Negative

emoji_wrapped_present		Positive
emoji_kissing_face_with_smiling_eyes	Yes	Positive
emoji_dizzy_face	Yes	Negative
emoji_heart_decoration		Positive
emoji_girl	Yes	
emoji_face_with_open_mouth	Yes	Ambigusous
emoji_frowning_face_with_open_mouth	Yes	Negative
emoji_hushed_face	Yes	Negative
emoji_bride_with_veil	Yes	
emoji_weary_cat_face		Negative
emoji_mouth	Yes	
emoji_swimmer	Yes	
emoji_white_left_pointing_index	Yes	
emoji_family	Yes	Ambiguous
emoji_man_with_turban	Yes	
emoji_person_frowning	Yes	Negative
emoji_haircut	Yes	
emoji_ear	Yes	
emoji_grinning_cat_face_with_smiling_eyes		Positive

Appendix B: Coding Guidance for tweet classification

Justice related

Relating to the U.S. criminal justice system. This system involves:

- investigation of criminal conduct
- arrests
- evidence gathered
- charges brought
- defenses raised
- trials conducted
- sentences rendered
- and punishment carried out

It includes

- Actions of actors in this system (e.g., police, prosecutors, etc.). Includes public statements and opinions, as well as behaviors.
- Events or other activities involving parties in the system (includes elected officials that drive criminal justice and policing, like the Mayor),
- Policies, processes and procedures relating to the criminal justice system
- Responses by individuals or groups (e.g., community members, the public, activists) to these justice-related actions, events, policies, processes, or procedures. Includes statements and opinions, as well as behaviors.

Tweets referencing the actions, views or opinions of the police, prosecutors, judges, etc.; tweets describing the actions, views or opinions of protestors or community members with respect to Freddie Gray's (or other's) arrest or death in a law enforcement context; and tweets about the events, consequences or outcomes of riots or protests relating to justice.

Indicators include mention of police officers, the prosecutor (Marilyn Moseby), use of a justice-related hashtag; emoticons of police, upraised fists, etc. News and commentary on these events, and on the justice system more generally are relevant.

Appendix C: Empath

achievement
affection
aggression
air_travel
alcohol
ancient
anger
animal
anonymity
anticipation
appearance
art
attractive
banking
beach
beauty
blue_collar_job
body
breaking
business
car
celebration
cheerfulness
childish
children
cleaning
clothing
cold
college
communication
competing
computer
confusion
contentment
cooking
crime

dance
death
deception
disappointment
disgust
dispute
divine
domestic_work
dominant_heirarchical
dominant_personality
driving
eating
economics
emotional
envy
exasperation
exercise
exotic
fabric
family
farming
fashion
fear
feminine
fight
fire
friends
fun
furniture
gain
giving
government
hate
healing
health
hearing
help
heroic
hiking
hipster

home
horror
hygiene
independence
injury
internet
irritability
journalism
joy
kill
law
leader
legend
leisure
liquid
listen
love
lust
magic
masculine
medical_emergency
medieval
meeting
messaging
military
money
monster
morning
movement
music
musical
negative_emotion
neglect
negotiate
nervousness
night
noise
occupation
ocean
office

optimism
order
pain
party
payment
pet
philosophy
phone
plant
play
politeness
politics
poor
positive_emotion
power
pride
prison
programming
rage
reading
real_estate
religion
restaurant
ridicule
royalty
rural
sadness
sailing
school
science
sexual
shame
shape_and_size
ship
shopping
sleep
smell
social_media
sound
speaking

sports
stealing
strength
suffering
superhero
surprise
swearing_terms
swimming
sympathy
technology
terrorism
timidity
tool
torment
tourism
toy
traveling
trust
ugliness
urban
vacation
valuable
vehicle
violence
war
warmth
water
weakness
wealthy
weapon
weather
wedding
white_collar_job
work
worship
writing
youth
zest

Appendix D: Model comparisons at different thresholds

Baltimore comparison of agreement at different thresholds for significant variables. The number of times each variable was significant across thresholds is shown in the last column.

	GT3					GT2					GT4					
	Estim	Std. Err	z value	Pr(>|z|)	Sig	Estim	Std. Err	z value	Pr(>|z|)	Sig	Estim	Std. Er	z value	Pr(>|z|)	Sig	Cnt Si
(Intercept)	-4.237	0.242	-17.49	< 2e-16	***	-3.460	0.227	-15.27	< 2e-16	***	-4.615	0.258	-17.92	< 2e-16	***	3
count	0.653	0.209	3.118	0.0018	**	0.688	0.189	3.65	0.0003	***	0.763	0.230	3.325	0.0009	***	3
activity	0.106	0.011	9.785	< 2e-16	***	0.119	0.011	10.981	< 2e-16	***	0.098	0.011	8.76	< 2e-16	***	3
J_tags	0.254	0.057	4.426	0.00001	***	0.211	0.054	3.9	1E-04	***	0.228	0.061	3.761	0.0002	***	3
R_tags	-0.217	0.075	-2.881	0.0040	**	-0.119	0.071	-1.681	0.093	.	-0.241	0.080	-3.022	0.0025	**	2
i	-0.105	0.059	-1.773	0.0762	.	-0.136	0.055	-2.473	0.013	*	-0.038	0.063	-0.607	0.5440		1
we	-0.071	0.062	-1.157	0.2473		-0.079	0.057	-1.368	0.171		-0.151	0.066	-2.273	0.0230	*	1
you	-0.233	0.063	-3.69	0.0002	***	-0.206	0.059	-3.521	0.000	***	-0.244	0.068	-3.599	0.0003	***	3
fight	-0.133	0.100	-1.336	0.1817		-0.216	0.094	-2.298	0.022	*	-0.126	0.106	-1.185	0.2360		1
posemo	-0.120	0.084	-1.423	0.1546		-0.167	0.079	-2.108	0.035	*	-0.156	0.090	-1.737	0.0824	.	1
kill	0.286	0.116	2.457	0.0140	*	0.247	0.110	2.256	0.024	*	0.375	0.123	3.051	0.0023	**	3
horror	0.318	0.145	2.187	0.0288	*	0.336	0.140	2.399	0.016	*	0.270	0.152	1.781	0.0749	.	2
exasp	-0.554	0.276	-2.008	0.0447	*	-0.688	0.270	-2.551	0.011	*	-0.582	0.282	-2.066	0.0388	*	3
suffer	-0.185	0.109	-1.69	0.0910	.	-0.123	0.102	-1.202	0.229		-0.282	0.117	-2.409	0.0160	*	1
negemo	-0.191	0.085	-2.247	0.0246	*	-0.144	0.078	-1.846	0.065	.	-0.156	0.091	-1.702	0.0887	.	1
law	0.150	0.103	1.454	0.1459		0.107	0.099	1.075	0.282		0.303	0.106	2.854	0.0043	**	1
violence	0.222	0.106	2.086	0.0370	*	0.167	0.099	1.685	0.092	.	0.244	0.114	2.14	0.0323	*	2
weapon	-0.173	0.086	-2.017	0.0437	*	-0.139	0.081	-1.711	0.087	.	-0.214	0.090	-2.366	0.0180	*	2
pic_parade	0.575	0.274	2.099	0.0359	*	0.604	0.238	2.537	0.011	*	0.466	0.306	1.524	0.1274		2
Betweenness	-7.528	3.500	-2.151	0.0315	*	-4.573	3.285	-1.392	0.164		-6.678	3.712	-1.799	0.0720	.	1
OutTieCount_DirectMention	0.271	0.068	4.012	0.0001	***	0.270	0.065	4.161	0.000	***	0.231	0.071	3.269	0.0011	**	3
OutTieCount_ReplyTo	-0.207	0.118	-1.751	0.07998	.	-0.316	0.115	-2.742	0.006	**	-0.131	0.122	-1.072	0.2835		1

favourites _count	-0.004	0.001	-2.781	0.00541	**	-0.004	0.001	-3.314	0.001	***	-0.005	0.002	-3.172	0.0015	**	3
friends_co unt	0.007	0.003	2.279	0.02269	*	0.006	0.003	2.264	0.024	*	0.007	0.004	1.898	0.0577	.	2
statuses_c ount	0.003	0.001	5.306	#######	***	0.003	0.001	4.996	0.000	***	0.002	0.001	3.538	0.0004	***	3
TFC_prof_i mage_cha nge	-0.297	0.153	-1.937	0.05278	.	-0.319	0.147	-2.171	0.030	*	-0.227	0.161	-1.414	0.1575		1

Baltimore model details at three thresholds

	GT3					GT2					GT4			
	Estim	Std. Err	z value	Pr(>|z|)	Sig	Estim	Std. Err	z value	Pr(>|z|)	Sig	Estim	Std. Er	z value	Pr(>|z|)
(Intercept)	-4.2374	0.2423	-17.487	< 2e-16	***	-3.4603	0.2267	-15.265	< 2e-16	***	-4.6148	0.2576	-17.918	< 2e-16
count	0.6525	0.2093	3.118	0.0018	**	0.6885	0.1886	3.65	0.0003	***	0.7634	0.2296	3.325	0.0009
activity	0.1060	0.0108	9.785	< 2e-16	***	0.1190	0.0108	10.981	< 2e-16	***	0.0976	0.0111	8.76	< 2e-16
FG_tags	0.0352	0.0573	0.614	0.5392		0.0362	0.0536	0.675	0.4998		0.0075	0.0611	0.123	0.9024
J_tags	0.2537	0.0573	4.426	9.6E-06	***	0.2110	0.0541	3.9	9.6E-05	***	0.2278	0.0606	3.761	0.0002
R_tags	-0.2171	0.0754	-2.881	0.0040	**	-0.1188	0.0707	-1.681	0.0928	.	-0.2414	0.0799	-3.022	0.0025
pos_emoji	-0.1605	0.1651	-0.972	0.3311		-0.2824	0.1573	-1.795	0.0726	.	-0.1191	0.1781	-0.669	0.5035
person_emoji	-0.0006	0.1478	-0.004	0.9966		0.1480	0.1434	1.032	0.3018		-0.2471	0.1569	-1.575	0.1152
neg_emoji	-0.1514	0.1557	-0.972	0.3309		-0.0638	0.1497	-0.426	0.6702		-0.1150	0.1687	-0.681	0.4956
i	-0.1048	0.0591	-1.773	0.0762	.	-0.1362	0.0551	-2.473	0.0134	*	-0.0383	0.0630	-0.607	0.5440
we	-0.0714	0.0617	-1.157	0.2473		-0.0785	0.0574	-1.368	0.1712		-0.1511	0.0665	-2.273	0.0230
you	-0.2329	0.0631	-3.69	0.0002	***	-0.2062	0.0585	-3.521	0.0004	***	-0.2442	0.0678	-3.599	0.0003
they	-0.0539	0.0660	-0.818	0.4134		-0.0204	0.0611	-0.334	0.7381		-0.0613	0.0707	-0.868	0.3855
aggr	-0.1458	0.1100	-1.325	0.1851		-0.0604	0.1036	-0.583	0.5596		-0.0844	0.1156	-0.73	0.4652
fight	-0.1333	0.0998	-1.336	0.1817		-0.2157	0.0939	-2.298	0.0216	*	-0.1256	0.1060	-1.185	0.2360
posemo	-0.1201	0.0844	-1.423	0.1546		-0.1675	0.0795	-2.108	0.0351	*	-0.1563	0.0900	-1.737	0.0824
kill	0.2859	0.1163	2.457	0.0140	*	0.2471	0.1096	2.256	0.0241	*	0.3747	0.1228	3.051	0.0023
death	0.0348	0.1048	0.332	0.7396		0.0518	0.0987	0.525	0.5999		-0.0071	0.1117	-0.064	0.9491
symp	0.3672	0.2295	1.6	0.1096		0.2776	0.2203	1.26	0.2076		0.2855	0.2424	1.178	0.2390
achiev	0.1974	0.1270	1.554	0.1201		0.2234	0.1218	1.835	0.0666	.	0.2353	0.1322	1.779	0.0752
religion	0.0287	0.1276	0.225	0.8219		0.0038	0.1230	0.031	0.9752		-0.1639	0.1347	-1.217	0.2235
horror	0.3181	0.1455	2.187	0.0288	*	0.3362	0.1402	2.399	0.0165	*	0.2701	0.1517	1.781	0.0749
order_ave	-0.0453	0.1075	-0.422	0.6734		0.0158	0.1033	0.153	0.8783		-0.0367	0.1112	-0.33	0.7414
shame	-0.0188	0.1151	-0.163	0.8702		-0.0143	0.1094	-0.13	0.8963		0.0427	0.1207	0.353	0.7238
exasp	-0.5542	0.2760	-2.008	0.0447	*	-0.6885	0.2699	-2.551	0.0107	*	-0.5821	0.2817	-2.066	0.0388
comm	0.0120	0.0971	0.124	0.9017		-0.0587	0.0917	-0.64	0.5220		0.0865	0.1028	0.841	0.4002
crime	-0.0725	0.1009	-0.718	0.4725		-0.0465	0.0933	-0.499	0.6179		-0.0625	0.1078	-0.58	0.5620
prison	-0.0228	0.0989	-0.231	0.8172		-0.0525	0.0938	-0.56	0.5756		-0.0539	0.1039	-0.519	0.6040
dispute	-0.0805	0.0764	-1.054	0.2918		-0.0636	0.0712	-0.894	0.3712		-0.0792	0.0814	-0.973	0.3306
speak	0.0755	0.0932	0.811	0.4176		0.1009	0.0866	1.165	0.2440		0.0536	0.0999	0.536	0.5917
suffer	-0.1849	0.1094	-1.69	0.0910	.	-0.1226	0.1020	-1.202	0.2293		-0.2820	0.1171	-2.409	0.0160
steal	0.1640	0.1049	1.564	0.1177		0.0993	0.0989	1.004	0.3152		0.1542	0.1109	1.391	0.1643
negemo	-0.1907	0.0849	-2.247	0.0246	*	-0.1435	0.0778	-1.846	0.0650	.	-0.1557	0.0915	-1.702	0.0887
pain	-0.0675	0.1081	-0.625	0.5321		0.0234	0.1014	0.231	0.8172		-0.1255	0.1143	-1.098	0.2720
listen	0.0033	0.0936	0.036	0.9716		0.1407	0.0879	1.601	0.1093		-0.1383	0.0999	-1.385	0.1662
law	0.1497	0.1029	1.454	0.1459		0.1066	0.0992	1.075	0.2824		0.3032	0.1062	2.854	0.0043
violence	0.2218	0.1063	2.086	0.0370	*	0.1666	0.0988	1.685	0.0919	.	0.2440	0.1140	2.14	0.0323
weapon	-0.1729	0.0857	-2.017	0.0437	*	-0.1391	0.0813	-1.711	0.0871	.	-0.2139	0.0904	-2.366	0.0180
pic_people	0.1239	0.3647	0.34	0.7341		0.1466	0.3100	0.473	0.6364		0.5544	0.3939	1.407	0.1593
pic_riot	0.1161	0.3156	0.368	0.7129		-0.0930	0.2755	-0.338	0.7357		0.2208	0.3486	0.633	0.5266
pic_protest	-0.1867	0.3558	-0.525	0.5998		-0.2455	0.3055	-0.804	0.4216		-0.6449	0.3843	-1.678	0.0933
pic_parade	0.5745	0.2738	2.099	0.0359	*	0.6036	0.2379	2.537	0.0112	*	0.4662	0.3058	1.524	0.1274
pic_police	-0.3206	0.2775	-1.155	0.2480		-0.2248	0.2442	-0.92	0.3575		-0.2736	0.3104	-0.881	0.3781
Betweenness	-7.5284	3.5003	-2.151	0.0315	*	-4.5729	3.2853	-1.392	0.1640		-6.6779	3.7123	-1.799	0.0720

Degree	-0.1845	0.1401	-1.317	0.1879		-0.2410	0.1363	-1.768	0.0771	.	-0.2012	0.1448	-1.389	0.1647
InTieCount_DirectMention	0.0987	0.0886	1.114	0.2653		0.0532	0.0876	0.608	0.5433		0.1305	0.0898	1.453	0.1462
InTieCount_ReplyTo	0.0586	0.1848	0.317	0.7511		0.0992	0.1814	0.547	0.5843		-0.0919	0.1907	-0.482	0.6298
InTieCount_Retweet	0.1076	0.0763	1.409	0.1589		0.1472	0.0760	1.937	0.0528	.	0.1007	0.0782	1.287	0.1982
OutTieCount_DirectMention	0.2714	0.0676	4.012	0.0001	***	0.2697	0.0648	4.161	0.00003	***	0.2311	0.0707	3.269	0.0011
OutTieCount_ReplyTo	-0.2070	0.1182	-1.751	0.07998	.	-0.3160	0.1152	-2.742	0.0061	**	-0.1307	0.1219	-1.072	0.2835
OutTieCount_Retweet	0.0785	0.0853	0.92	0.35758		0.0701	0.0818	0.857	0.39136		0.1023	0.0890	1.149	0.2504
favourites_count	-0.0040	0.0014	-2.781	0.00541	**	-0.0044	0.0013	-3.314	0.00092	***	-0.0049	0.0015	-3.172	0.0015
followers_count	-0.0011	0.0023	-0.467	0.64041		0.0000	0.0020	-0.018	0.98543		-0.0014	0.0026	-0.531	0.5955
friends_count	0.0071	0.0031	2.279	0.02269	*	0.0062	0.0028	2.264	0.02361	*	0.0067	0.0035	1.898	0.0577
statuses_count	0.0034	0.0006	5.306	#######	***	0.0029	0.0006	4.996	#######	***	0.0025	0.0007	3.538	0.0004
Closeness_orig	0.0001	0.0001	0.972	0.33097		0.0000	0.0001	0.302	0.76251		0.0001	0.0001	1.202	0.2295
TFC_prof_image_change	-0.2968	0.1532	-1.937	0.05278	.	-0.3187	0.1468	-2.171	0.02994	*	-0.2271	0.1607	-1.414	0.1575
TFC_descrip_change	-0.1069	0.1814	-0.589	0.55564		-0.0396	0.1718	-0.23	0.81787		-0.0319	0.1898	-0.168	0.8667
TFC_inMC	-0.0981	0.3534	-0.277	0.78145		0.0694	0.3068	0.226	0.82103		-0.33862	0.4052	-0.836	0.4034

Cleveland comparison of agreement at different thresholds for significant variables. The number of times each variable was significant across thresholds is shown in the last column

	Estim	Std. Er	z value	Pr(>|z|)	sig	Estim	Std. Err	z value	Pr(>|z|)	sig	Estim	Std. Err	z value	Pr(>|z|)	sig	Cnt Sig
(Intercept)	-4.8009	0.6005	-7.995	1.29E-15	***	-4.1218	0.5700	-7.231	4.80E-13	***	-4.9440	0.6583	-7.511	5.89E-14	***	3
count	-1.6333	0.4149	-3.936	8.27E-05	***	-1.7014	0.3810	-4.465	7.99E-06	***	-1.9489	0.4801	-4.06	4.91E-05	***	3
activity.x	0.1362	0.0421	3.238	0.0012	**	0.1622	0.0420	3.859	0.0001	***	0.1796	0.0446	4.03	5.58E-05	***	3
J_tags	-0.5394	0.1308	-4.124	3.73E-05	***	-0.3973	0.1191	-3.337	0.0008	***	-0.4522	0.1464	-3.088	0.0020	**	3
achiev	-0.9566	0.3887	-2.461	0.0139	*	-1.0263	0.3637	-2.822	0.0048	**	-0.9036	0.4332	-2.086	0.0370	*	3
horror	-0.6624	0.4408	-1.503	0.1329		-0.9769	0.4143	-2.358	0.0184	*	0.2589	0.4685	0.553	0.5805		1
pic_parade	-1.3079	0.5946	-2.2	0.0278	*	-0.5156	0.5608	-0.919	0.3580		-0.4155	0.7408	-0.561	0.5749		1
pic_police	1.1761	0.4842	2.429	0.0151	*	0.7754	0.4524	1.714	0.0865	.	1.7539	0.5592	3.137	0.0017	**	2
Betweenness	7.8119	4.2905	1.821	0.0687	.	15.4889	4.0421	3.832	0.0001	***	7.7776	4.8124	1.616	0.1061		1
Degree	1.1680	0.3631	3.217	0.0013	**	1.1758	0.3597	3.268	0.0011	**	0.9919	0.3822	2.595	0.0095	**	3
OutTieCount_DirectMention	0.3962	0.1425	2.78	0.0054	**	0.2312	0.1396	1.656	0.0976	.	0.6291	0.1540	4.084	0.0000	***	2
OutTieCount_ReplyTo	0.5932	0.2465	2.407	0.0161	*	0.5178	0.2566	2.018	0.0436	*	0.3917	0.2569	1.525	0.1272		2
OutTieCount_Retweet	0.9759	0.2016	4.84	0.0000	***	0.7451	0.1902	3.917	0.0001	***	1.1949	0.2208	5.411	0.0000	***	3
favourites_count	-0.0078	0.0028	-2.8	0.0051	**	-0.0080	0.0026	-3.138	0.0017	**	-0.0061	0.0031	-1.968	0.0491	*	3
followers_count	0.0042	0.0022	1.902	0.0572	.	0.0019	0.0023	0.819	0.4129		0.0050	0.0024	2.098	0.0359	*	1
statuses_count	0.0050	0.0012	4.119	0.0000	***	0.0048	0.0012	4.161	0.0000	***	0.0034	0.0013	2.543	0.0110	*	3
Closeness_orig	0.0004	0.0001	3.137	0.0017	**	0.0005	0.0001	4.284	0.0000	***	0.0005	0.0002	3.245	0.0012	**	3
TFC_inMC	-2.4989	0.7263	-3.441	0.0006	***	-2.7874	0.6363	-4.38	0.0000	***	-3.2780	0.8792	-3.728	0.0002	***	3

Cleveland model details at three thresholds

	Estim	Std. Er	z value	Pr(>|z|)	sig	Estim	Std. Err	z value	Pr(>|z|)	sig	Estim	Std. Err	z value	Pr(>|z
(Intercept)	-4.8009	0.6005	-7.995	1.29E-15	***	-4.1218	0.5700	-7.231	4.80E-13	***	-4.9440	0.6583	-7.511	5.89 1
count	-1.6333	0.4149	-3.936	8.27E-05	***	-1.7014	0.3810	-4.465	7.99E-06	***	-1.9489	0.4801	-4.06	4.91 0
activity.x	0.1362	0.0421	3.238	0.0012	**	0.1622	0.0420	3.859	0.0001	***	0.1796	0.0446	4.03	5.58 0
TR_tags	-0.1855	0.1878	-0.988	0.3232		-0.1058	0.1834	-0.577	0.5641		-0.1908	0.2071	-0.921	0.357
J_tags	-0.5394	0.1308	-4.124	3.73E-05	***	-0.3973	0.1191	-3.337	0.0008	***	-0.4522	0.1464	-3.088	0.002
pos_emoji	-0.1075	0.6564	-0.164	0.8699		-0.9308	0.5848	-1.592	0.1115		-0.2535	0.7323	-0.346	0.729
person_emoji	-0.4911	0.7184	-0.684	0.4943		0.3703	0.5508	0.672	0.5014		-0.4190	0.7763	-0.54	0.589
neg_emoji	-0.4838	0.7098	-0.682	0.4954		0.5685	0.5880	0.967	0.3336		-0.1502	0.7839	-0.192	0.848
i	0.0357	0.1361	0.263	0.7929		-0.1076	0.1246	-0.864	0.3878		0.0082	0.1562	0.052	0.958
we	-0.0210	0.1577	-0.133	0.8939		0.1262	0.1459	0.865	0.3872		0.0928	0.1723	0.539	0.590
you	-0.2268	0.1512	-1.5	0.1336		-0.2455	0.1389	-1.768	0.0771	.	-0.1819	0.1671	-1.089	0.276
they	-0.0195	0.1750	-0.112	0.9112		-0.0823	0.1604	-0.513	0.6081		0.0047	0.1970	0.024	0.981
aggr	0.1453	0.2703	0.537	0.5909		0.3357	0.2566	1.308	0.1909		-0.4568	0.3076	-1.485	0.137
fight	0.3175	0.2356	1.348	0.1777		0.0726	0.2229	0.326	0.7445		0.3105	0.2631	1.18	0.238
posemo	-0.2651	0.2015	-1.315	0.1884		-0.0894	0.1847	-0.484	0.6281		0.0220	0.2202	0.1	0.920
kill	-0.1207	0.2698	-0.447	0.6547		0.1272	0.2539	0.501	0.6165		0.0790	0.3018	0.262	0.793
death	0.0771	0.2552	0.302	0.7626		-0.0791	0.2401	-0.329	0.7418		0.1786	0.2852	0.626	0.531
symp	0.7056	0.5373	1.313	0.1891		0.3834	0.5257	0.729	0.4658		0.7075	0.5997	1.18	0.238
achiev	-0.9566	0.3887	-2.461	0.0139	*	-1.0263	0.3637	-2.822	0.0048	**	-0.9036	0.4332	-2.086	0.037
religion	-0.1341	0.3991	-0.336	0.7370		-0.4910	0.3858	-1.273	0.2031		-0.3268	0.4514	-0.724	0.469
horror	-0.6624	0.4408	-1.503	0.1329		-0.9769	0.4143	-2.358	0.0184	*	0.2589	0.4685	0.553	0.580
order_ave	0.0220	0.2852	0.077	0.9386		0.2460	0.2593	0.949	0.3428		-0.2204	0.3297	-0.669	0.503
shame	0.3604	0.2807	1.284	0.1992		0.4922	0.2655	1.854	0.0638	.	0.1745	0.3179	0.549	0.582
exasp	0.8611	0.5825	1.478	0.1393		0.8985	0.5801	1.549	0.1214		-0.0918	0.7258	-0.127	0.899
comm	-0.3056	0.2435	-1.255	0.2095		-0.2585	0.2266	-1.141	0.2539		-0.2929	0.2729	-1.073	0.283
crime	0.0045	0.2092	0.022	0.9828		-0.1014	0.1915	-0.53	0.5964		0.2412	0.2350	1.026	0.304
prison	0.0341	0.2466	0.138	0.8902		0.0403	0.2283	0.176	0.8601		-0.1703	0.2769	-0.615	0.538
dispute	0.1449	0.1694	0.856	0.3922		0.1238	0.1548	0.8	0.4239		0.0818	0.1904	0.429	0.667
speak	0.3881	0.2001	1.939	0.0525	.	0.2007	0.1869	1.074	0.2828		0.1809	0.2254	0.803	0.422
suffer	-0.0248	0.2640	-0.094	0.9253		-0.0857	0.2537	-0.338	0.7355		-0.2900	0.2923	-0.992	0.321
steal	0.1135	0.2320	0.489	0.6247		0.0987	0.2128	0.464	0.6429		-0.1618	0.2619	-0.618	0.536
negemo	0.0070	0.1676	0.041	0.9669		-0.0471	0.1536	-0.307	0.7591		-0.0323	0.1898	-0.17	0.865

pain	-0.1780	0.2699	-0.659	0.5097		-0.2151	0.2590	-0.831	0.4062		0.0491	0.2944	0.167	0.867
listen	-0.0262	0.2329	-0.113	0.9103		0.0195	0.2185	0.089	0.9290		-0.2636	0.2617	-1.007	0.313
law	0.0218	0.2502	0.087	0.9304		-0.0306	0.2336	-0.131	0.8958		0.0814	0.2807	0.29	0.771
violence	0.2179	0.2593	0.84	0.4007		-0.0516	0.2451	-0.211	0.8332		0.5269	0.2899	1.818	0.069
weapon	-0.1484	0.1721	-0.863	0.3884		-0.1660	0.1600	-1.037	0.2995		-0.1547	0.1926	-0.803	0.422
pic_people	0.2657	0.5984	0.444	0.6570		-0.0576	0.5438	-0.106	0.9156		0.3255	0.7266	0.448	0.654
pic_riot	0.4957	0.6472	0.766	0.4438		0.9192	0.6037	1.523	0.1279		-0.1712	0.8270	-0.207	0.836
pic_protest	0.0214	0.5674	0.038	0.9699		-0.3372	0.5303	-0.636	0.5248		-0.7476	0.7677	-0.974	0.330
pic_parade	-1.3079	0.5946	-2.2	0.0278	*	-0.5156	0.5608	-0.919	0.3580		-0.4155	0.7408	-0.561	0.574
pic_police	1.1761	0.4842	2.429	0.0151	*	0.7754	0.4524	1.714	0.0865	.	1.7539	0.5592	3.137	0.001
Betweenness	7.8119	4.2905	1.821	0.0687	.	15.4889	4.0421	3.832	0.0001	***	7.7776	4.8124	1.616	0.106
Degree	1.1680	0.3631	3.217	0.0013	**	1.1758	0.3597	3.268	0.0011	**	0.9919	0.3822	2.595	0.009
InTieCount_DirectMention	-0.1071	0.1952	-0.548	0.5835		0.1262	0.2063	0.612	0.5408		-0.0879	0.2012	-0.437	0.662
InTieCount_ReplyTo	0.3792	0.3890	0.975	0.3297		0.3439	0.4512	0.762	0.4459		0.5988	0.4028	1.487	0.137
InTieCount_Retweet	0.3324	0.2175	1.529	0.1264		0.3577	0.2238	1.598	0.1100		0.3615	0.2230	1.621	0.104
OutTieCount_DirectMention	0.3962	0.1425	2.78	0.0054	**	0.2312	0.1396	1.656	0.0976	.	0.6291	0.1540	4.084	0.000
OutTieCount_ReplyTo	0.5932	0.2465	2.407	0.0161	*	0.5178	0.2566	2.018	0.0436	*	0.3917	0.2569	1.525	0.127
OutTieCount_Retweet	0.9759	0.2016	4.84	0.0000	***	0.7451	0.1902	3.917	0.0001	***	1.1949	0.2208	5.411	0.000
favourites_count	-0.0078	0.0028	-2.8	0.0051	**	-0.0080	0.0026	-3.138	0.0017	**	-0.0061	0.0031	-1.968	0.049
followers_count	0.0042	0.0022	1.902	0.0572	.	0.0019	0.0023	0.819	0.4129		0.0050	0.0024	2.098	0.035
friends_count	0.0027	0.0050	0.548	0.5835		0.0076	0.0048	1.592	0.1113		0.0012	0.0056	0.224	0.822
statuses_count	0.0050	0.0012	4.119	0.0000	***	0.0048	0.0012	4.161	0.0000	***	0.0034	0.0013	2.543	0.011
Closeness_orig	0.0004	0.0001	3.137	0.0017	**	0.0005	0.0001	4.284	0.0000	***	0.0005	0.0002	3.245	0.001
TFC_prof_image_change	0.0744	0.6075	0.123	0.9025		-0.5889	0.6072	-0.97	0.3321		0.2582	0.6197	0.417	0.676
TFC_descrip_change	-0.1258	0.6671	-0.189	0.8505		1.0380	0.6761	1.535	0.1247		-0.1643	0.6907	-0.238	0.812
TFC_inMC	-2.4989	0.7263	-3.441	0.0006	***	-2.7874	0.6363	-4.38	0.0000	***	-3.2780	0.8792	-3.728	0.000

Appendix E: Hashtags identified for Baltimore and Cleveland

Hashtags used for calculating hashtag features for Baltimore and Cleveland

Freddie Gray

asongforfreddiegray
biggerthanfreddiegray
frddiegray
freddie
freddieg
freddiegr
freddiegra
freddiegray
freddiegrayfuneral
freddiegraymovement
freddiegrayprotest
freddiegrayresteasy
freddiegrays
freddiegraysfamily
freddiegrayw
freddiegrey
freddygray
freediegray
iamfreddiegray
juctice4freddie
jusiceforfreddie
justforfreddie
justice4freddi
justice4freddie
justice4freddieg
justice4freddiegray
justice4freddievideo
justice4freddygray
justiceforfreddie
justiceforfreddiegray
justiceforfreddiegrey
justiceforfreddy
justicesforfreddiegray
juticeforfreddiegrey
longlivefreddie
longlivefreddieg
longlivefreddiegray
mayday4freddiegray
muslims4gray
prayforfreddie
rallyforfreddygray
restinpowerreddiegray
ripfreddie
ripfreddieg
ripfreddiegray
ripfreddygray
ripfreddygrey
ripfreediegray
studentsforfreddiegray

Riot

bailtimoreriots
balitimoreriots
balitmoreriots
baltimoreburning
baltimorecityriot
baltimorelootcrew
baltimorepoliceriot
baltimoreriot
baltimorerioters
baltimoreriotpolice
baltimoreriots
baltimroeriots
baltmoreriots
dontburnbaltimore
fewstoreslooted
giveevanriotgear
looting
mondawminriot
noriots
notariot
notthebaltimoreriots
riot
riotcantstopus
rioters
riotfest
rioting
riotmap
riotorprotest
riots
stopthelooting
theriot

Baltimore justice

21stcenturypolicingtaskforce
24hourcurfew
aacopd
accidentlaw
aftercurfew
allblacklivesmatter
alllifesmatter
alllivesmatter
alllivesmatters
allpoclivesmatter
anarchyisnotaprotest
antipolicebrutality
arrestpolice
badcop
balitmoreuprising
balrimoreuprising
baltiimoreuprising
baltimoreaftercurfew
baltimorecitypolice
baltimorecops
baltimorecrime
baltimorecurfew
baltimorejusticefund
baltimorelivesmatter
baltimorepolice
baltimorepolicedepartment
baltimorepoliceriot
baltimoreprotes
baltimoreprotest
baltimoreprotests
baltimoreriotpolice
baltimoreuprise
baltimoreuprisi
baltimoreuprising
baltimoreuprisings
bcopd
beatthecurfew
bkacklivesmatter
blackfemalelivesmatter
blacklawyer
blacklivematter
blackliveshavealwaysmattered
blacklivesmatter
blacklivesmatternyc
blacklivesmatters
blacklivesstillmatter
blackliveswillalwaysmatter
blackqueerlivesmatter
blackslivesmatter
blacktranslivesmatter
blacktranswomenmatter
blackwomenlivesmatter
blackwomenslivesmatter
bluelivesmatter
bmorecurfew
bmoreuprising
breakingcurfew
breakthecurfew
brownlivesmatter
cancelthecurfew
communitypolicing
controlcriminalsnotnra
cop
copaganda
copblock
copdoc
copforaday
cophater
cops
copsarerunningaway
copsbiteballs
copsdontdotime
copskill
copslivesmatter
copswillbiteyourballs
copwatch
crime
crimedoesntpay
criminal
criminalcharges
criminaldamage
criminaljustice
criminallaw
crimingwhilewhite
curfew
curfewlifted
curfewover
curfewproblems
curfewreax
curfewsucks

dcrising
deathbycop
decriminalizeblack
disabledblackwomenmatter
domesticlaw
endpolicebrutality
endpoliceterrorism
endthecurfew
ferguson
filmthepolice
fixthepolice
fkkthepolice
freddiegrayprotest
fuckdalaw
fuckdalaww
fuckdapolice
fuckpolice
fuckthecurfew
fuckthepolice
fukacurfew
fukyacurfew
goodcops
hugacop
juctice4freddie
jusiceforfreddie
justdg
justice
justice4freddi
justice4freddie
justice4freddieg
justice4freddiegray
justice4freddievideo
justice4freddygray
justiceforbaltimore
justiceforfreddie
justiceforfreddiegray
justiceforfreddiegrey
justiceforfreddy
justiceformartese
justiceformikebrown
justiceforrekia
justicefortamirrice
justiceforthem
justicehasbeenserved
justicesforfreddiegray
juvenilejustice
kidsnotcriminals
killedbypolice
killercops
law
lawabidingcitizen
lawenforcement
lawfirms
lawless
lawschool
lawyer
lawyerintraining
lawyers
liftthecurfew
march2justice
mdlawyers
militarizationofpolice
morebaltimorecopstories
muslimlivesmatter
mylifematters
nationalpoliceweek
nativelivesmatter
nocops
nocurfew
nojustice
nojusticeforcopsnopeace
nojusticenopeace
nojusticenopeacenoracistpolice
nopolicestate
noracistpolice
notallcops
obeythecurfew
osdontsupportpolicebrutality
ourlivesmatter
peacefulprotest
peacefulprotestors
peacefulprotests
police
policebodycams
policebrutality
policebrutalityisatanalltimehigh
policekillings
policeliesmatter
policelivesmatter
policememorialweek
policemilitarization
polices

policestate
policethepolice
policeunitytour
policeviolence
policeweek
policeweekm
policeweekmatters
policing
prelaw
protest
protesters
protestforjustice
protestpeacefully
protests
rallyforfreddygray
respectingthecurfew
riotorprotest
ruleoflaw
sayhername
selectivecurfew
smashthesecops
specialcriminalcourt
stopkillercops
stoppolice
stoppolicebrutality
stopthecurfew
stopwhiteonwhitecrime
studentsforfreddiegray
translivesmatter
tulsapolice
tyronewest
uncuffourcops
walterscott
wefight4justice
wenotallcriminals
whereisjustice
whitelivesmatter
whiteonwhitecrime
youarenotagoodcop

Baltimore general

baltimorestrong
baltimoreweareprayiing4u
bmorebloc
bmorestrong
bmoreunited
charmcitystrong
onebaltimore
pray4baltimore
pray4bmore
prayer4baltimore
prayersforbaltimore
prayforbaltim
prayforbaltimor
prayforbaltimore
prayforbaltinore
prayforbmore
prayformycity
prayforourpeopleandcity
prayforpeaceinbaltimore
praying4baltimore
prayingforbaltimore
prayingformycity
standstrongbaltimore
weareonebaltimore

Cleveland hashtags

Tamir Rice

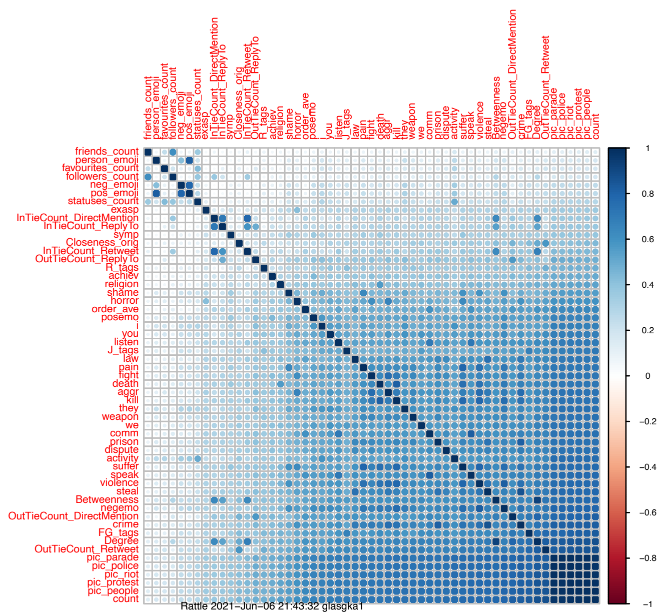
justice4tamir
justicefortamirrice
riptamirrice
tamir
tamirrice

Cleveland justice

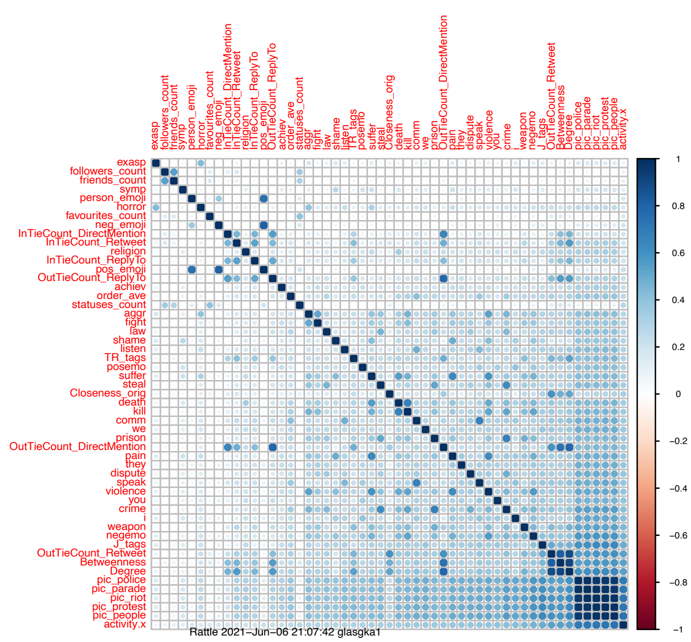
alllivesmatter
blacklivesmatter
bluelivesmatter
chapelhillshooting
clepolice
clevelandpolice
crimingwhilewhite
darrenwilson
ericgarner
ferguson
ferguson2cle
fergusondecision
icantbreathe
justice
justice4all
justice4mikebrown
justice4tamir
justiceforericgarner
justiceformikebrown
justicefortamirrice
justicefortanishaanderson
michaelbrown
mikebrown
nojusticenopeace
police
policebrutality
policelivesmatter
ripericgarner
riptamirrice
seaofblue
tanishaanderson
thinblueline

Appendix F: Correlations for Baltimore and Cleveland variables used in logistic regression

Correlation Buser_sqrt_close.csv using Pearson



Correlation Cuser_sqrt_close.csv using Pearson



Glossary

If needed.

Bibliography

Bennett, W. L., & Segerberg, A. (2011). Digital Media and the Personalization of Collective Action. *Information, Communication & Society*, *14*(6), 770–799. https://doi.org/10.1080/1369118X.2011.579141

Bennett, W. L., & Segerberg, A. (2012). The Logic of Connective Action. *Information, Communication & Society*, *15*(5), 739–768. https://doi.org/10.1080/1369118X.2012.670661

Bodin, Ö., & Crona, B. I. (2009). The role of social networks in natural resource governance: What relational patterns make a difference? *Global Environmental Change*, *19*(3), 366–374. https://doi.org/10.1016/j.gloenvcha.2009.05.002

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*(7415), 295–298. https://doi.org/10.1038/nature11421

Bonilla, Y., & Rosa, J. (2015). #Ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the United States. *American Ethnologist*, *42*(1), 4–17. https://doi.org/10.1111/amet.12112

Borgatti, S. P., & Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. *Journal of Management*, *29*(6), 991.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, *323*(5916), 892–895. https://doi.org/10.1126/science.1165821

boyd, danah, & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

boyd, D., Golder, S., & Lotan, G. (2010). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences*, 1–10. https://doi.org/10.1109/HICSS.2010.412

Brown, G., & Pickerill, J. (2009). Space for emotion in the spaces of activism. *Emotion, Space and Society*, *2*(1), 24–35. https://doi.org/10.1016/j.emospa.2009.03.004

Bruns, A., & Burgess, J. E. (2011). The use of Twitter hashtags in the formation of ad hoc publics. *Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011*. http://eprints.qut.edu.au/46515

Cantijoch, M., Cutts, D., & Gibson, R. (2016). Moving Slowly up the Ladder of Political Engagement: A ‘Spill-over’ Model of Internet Participation. *The British Journal of Politics and International Relations*, *18*(1), 26–48. https://doi.org/10.1111/1467-856X.12067

Christakis, N. A., & Fowler, J. H. (2007). The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine*, *357*(4), 370–379. https://doi.org/10.1056/NEJMsa066082

Conley, A., & Moote, M. A. (2003). Evaluating collaborative natural resourcemanagement. *Society &Natural Resources*, *16*(5), 371–386.

Conover, M. D., Ferrara, E., Menczer, F., & Flammini, A. (2013). The Digital Evolution of Occupy Wall Street. *PLOS ONE*, *8*(5), e64679. https://doi.org/10.1371/journal.pone.0064679

Conover, M. D., Ratkiewicz, J., Francisco, M., Gon\ccalves, B., Flammini, A., & Menczer, F. (2011). Political polarization on twitter. *Proc. 5th Intl. Conference on Weblogs and Social Media*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2847/3275

Croeser, S., & Highfield, T. (2014). Occupy Oakland and #oo: Uses of Twitter within the Occupy movement. *First Monday*, *19*(3), 1–13.

Cropanzano, R. S., Ambrose, M. L., Blader, S. L., & Tyler, T. R. (2015). Relational Models of Procedural Justice. In R. S. Cropanzano & M. L. Ambrose (Eds.), *The Oxford Handbook of Justice in the Workplace*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199981410.013.16

Cropanzano, R. S., Ambrose, M. L., Masterson, S. S., & Tong, N. (2015). Justice Perception Formation in Social Settings. In R. S. Cropanzano & M. L. Ambrose (Eds.), *The Oxford Handbook of Justice in the Workplace*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199981410.013.13

De Choudhury, M., Jhaver, S., Sugar, B., & Weber, I. (2016). Social Media Participation in an Activist Movement for Racial Equality. *Tenth International*

AAAI Conference on Web and Social Media. http://www.munmund.net/pubs/BLM_ICWSM16.pdf

de Zuniga, H. G., Veenstra, A., Vraga, E., & Shah, D. (2010). Digital Democracy: Reimagining Pathways to Political Participation. *Journal of Information Technology & Politics*, *7*(1), 36–51. https://doi.org/10.1080/19331680903316742

Diani, M. (2008). Networks and Participation. In D. A. Snow, S. A. Soule, & H. Kriesi (Eds.), *The Blackwell companion to social movements* (pp. 339–360). John Wiley & Sons. https://books.google.com/books?hl=en&lr=&id=VYrPtQGrKkIC&oi=fnd&pg=PR5&dq=%22D.+Morris+and+Suzanne%22+%22Transnational+Processes+and%22+%22Collective+Identity,+Solidarity,+and%22+%22Diffusion+Processes+within+and+across%22+%22A.+Hunt+and+Robert+D.%22+%22Networks+and%22+%22della+Porta+and+Olivier%22+&ots=nrMJznj0Jc&sig=UhvrVUZpVzqMcIL-MI-C3LvdCjY

Diani, M. (2015). Social Movements, Networks and. *The Blackwell Encyclopedia of Sociology*. https://doi.org/10.1002/9781405165518.wbeoss162.pub2

Douglas, J. D. (1967). The Challenge of Crime in a Free Society: A Report by the President's Commission on Law Enforcement and Administration of Justice. *American Sociological Review*, *32*(4), 664. https://doi.org/10.2307/2091056

Dredze, M. (2012). How social media will change public health. *Intelligent Systems, IEEE*, *27*(4), 81–84.

Ellison, N. B., & boyd, D. M. (2013). *Sociality Through Social Network Sites*. https://doi.org/10.1093/oxfordhb/9780199589074.013.0008

Engel, R. S. (2005). Citizens' perceptions of distributive and procedural injustice during traffic stops with police. *Journal of Research in Crime and Delinquency*, *42*(4), 445–481.

Evans, M. P. (2013). Mapping social media networks in youth organizing. *Journal of Information, Information Technology, and Organizations*, *8*, 67–82.

Fast, E., Chen, B., & Bernstein, M. S. (2016). Empath: Understanding Topic Signals in Large-Scale Text. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 4647–4657. https://doi.org/10.1145/2858036.2858535

Freelon, D. (2014). Online Civic Activism: Where Does It Fit? *Policy & Internet*, *6*(2), 192–198. https://doi.org/10.1002/1944-2866.POI358

Freelon, D., McIlwain, C. D., & Clark, M. (2016). *Beyond the Hashtags: #Ferguson, #Blacklivesmatter, and the Online Struggle for Offline Justice* (SSRN Scholarly Paper ID 2747066). Social Science Research Network. https://doi.org/10.2139/ssrn.2747066

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

Gallagher, R. J., Reagan, A. J., Danforth, C. M., & Dodds, P. S. (2018). Divergent discourse between protests and counter-protests: #BlackLivesMatter and #AllLivesMatter. *PLOS ONE*, *13*(4), e0195644. https://doi.org/10.1371/journal.pone.0195644

Gifford, C., Rodriguez, P., Amundson, R., Awad, S., Chee, B., Lau, P., & Patrikar, A. (2015). Intelligent Fusion of CNN Ensembles. *ImageNet and MS COCO Visual Recognition Challenges Joint Workshop*.

Glasgow, K., Fink, C., & Boyd-Graber, J. (2014). *"Our Grief is Unspeakable": Automatically Measuring the Community Impact of a Tragedy*.

Glasgow, K., Tausczik, Y., & Vitak, J. (2020). *Beyond the# justice hashtags: Understanding digital participation after police-involved Deaths*.

Gleason, B. (2013). #Occupy Wall Street Exploring Informal Learning About a Social Movement on Twitter. *American Behavioral Scientist*, *57*(7), 966–982. https://doi.org/10.1177/0002764213479372

Goffman, E. & others. (1959). The presentation of self in everyday life. 1959. *Garden City, NY*, *259*.

Golbeck, J., Robles, C., & Turner, K. (2011). Predicting Personality with Social Media. *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, 253–262. https://doi.org/10.1145/1979742.1979614

Goleman, D. P. (1995). *Emotional intelligence: Why it can matter more than IQ for character, health and lifelong achievement*. New York: Bantam Books.

Gordon, E., Baldwin-Philippi, J., & Balestra, M. (2013). *Why We Engage: How Theories of Human Behavior Contribute to Our Understanding of Civic Engagement in a Digital Era* (SSRN Scholarly Paper ID 2343762). Social Science Research Network. http://papers.ssrn.com/abstract=2343762

Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*, *91*(3), 481–510.

Granovetter, M. S. (1973). The Strength of Weak Ties. *The American Journal of Sociology*, *78*(6), 1360–1380.

Hanusch, F., & Bruns, A. (2017). Journalistic Branding on Twitter. *Digital Journalism*, *5*(1), 26–43. https://doi.org/10.1080/21670811.2016.1152161

He, J., Hong, L., Frias-Martinez, V., & Torrens, P. (2015). Uncovering Social Media Reaction Pattern to Protest Events: A Spatiotemporal Dynamics Perspective of Ferguson Unrest. *SocInfo*, 67–81. http://www.vanessafriasmartinez.org/uploads/SocinfoPaper.pdf

Heverin, T., & Zach, L. (2010). *Microblogging for Crisis Communication: Examination of Twitter Use in Response to a 2009 Violent Crisis in the Seattle-Tacoma, Washington, Area*. ISCRAM. http://www.researchgate.net/profile/Lisl_Zach/publication/268361197_Microblogging_for_Crisis_Communication_Examination_of_Twitter_Use_in_Response_to_a_2009_Violent_Crisis_in_the_Seattle-Tacoma_Washington_Area/links/551a8b680cf244e9a4588838.pdf

Hong, L., Yang, W., Resnik, P., & Frias-Martinez, V. (2016). Uncovering Topic Dynamics of Social Media and News: The Case of Ferguson. *International Conference on Social Informatics*, 240–256. http://link.springer.com/chapter/10.1007/978-3-319-47880-7_15

Howard, P. N., Agarwal, S. D., & Hussain, M. M. (2011). When Do States Disconnect Their Digital Networks? Regime Responses to the Political Uses of Social Media. *The Communication Review*, *14*(3), 216–232. https://doi.org/10.1080/10714421.2011.597254

Howison, J., Crowston, K., & Wiggins, A. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal for the Association of Information Systems*.

Huberman, B. A., Romero, D. M., & Wu, F. (2008). *Social Networks that Matter: Twitter Under the Microscope* (SSRN Scholarly Paper ID 1313405). Social Science Research Network. https://papers.ssrn.com/abstract=1313405

Hurwitz, J., & Peffley, M. (2005). Explaining the Great Racial Divide: Perceptions of Fairness in the U.S. Criminal Justice System. *Journal of Politics*, *67*(3), 762–783. https://doi.org/10.1111/j.1468-2508.2005.00338.x

Jasper, J. M. (1998). The Emotions of Protest: Affective and Reactive Emotions In and Around Social Movements. *Sociological Forum*, *13*(3), 397–424. https://doi.org/10.1023/A:1022175308081

Juris, J. S. (2012). Reflections on #Occupy Everywhere: Social media, public space, and emerging logics of aggregation. *American Ethnologist*, *39*(2), 259–279. https://doi.org/10.1111/j.1548-1425.2012.01362.x

Justice Department Releases Report on Civil Rights Division's Pattern and Practice Police Reform Work. (2017, January 4). https://www.justice.gov/opa/pr/justice-department-releases-report-civil-rights-division-s-pattern-and-practice-police-reform

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Klandermans, B. (1996). *The social psychology of protest*.

Kochel, T. (2014). Views By St Louis County Residents Regarding the Police and Public Responses to the Shooting of Michael Brown in Ferguson, Missouri on August 9, 2014. *Reports*. http://opensiuc.lib.siu.edu/ccj_reports/2

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097–1105. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th International Conference on World Wide Web*, 591–600.

Lazer, D., Pentland, A. (Sandy), Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Life in the network: The coming age of computational social science. *Science (New York, N.Y.)*, *323*(5915), 721–723. https://doi.org/10.1126/science.1167742

Lee, J., & Kim, S. (2011). Exploring the Role of Social Networks in Affective Organizational Commitment: Network Centrality, Strength of Ties, and Structural Holes. *The American Review of Public Administration*, *41*(2), 205–223. https://doi.org/10.1177/0275074010373803

Leong, C., Pan, S. L., Bahri, S., & Fauzi, A. (2019). Social media empowerment in social movements: Power activation and power accrual in digital activism. *European Journal of Information Systems*, *28*(2), 173–204. https://doi.org/10.1080/0960085X.2018.1512944

Lind, E. A., Kray, L., & Thompson, L. (1998). The Social Construction of Injustice: Fairness Judgments in Response to Own and Others' Unfair Treatment by Authorities. *Organizational Behavior and Human Decision Processes*, *75*(1), 1–22. https://doi.org/10.1006/obhd.1998.2785

Loper, E., & Bird, S. (2002). NLTK: The natural language toolkit. *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics-Volume 1*, 63–70. http://dl.acm.org/citation.cfm?id=1118117

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., & others. (2011). The Arab Spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International Journal of Communication*, *5*, 31.

Mahrt, M., & Scharkow, M. (2013). The Value of Big Data in Digital Media Research. *Journal of Broadcasting & Electronic Media*, *57*(1), 20–33. https://doi.org/10.1080/08838151.2012.761700

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55–60. https://doi.org/10.3115/v1/P14-5010

Marsden, P. V. (1990). Network Data and Measurement. *Annual Review of Sociology*, *16*, 435–463.

Mcclurg, S. D. (2003). Social Networks and Political Participation: The Role of Social Interaction in Explaining Political Participation. *Political Research Quarterly*, *56*(4), 449–464. https://doi.org/10.1177/106591290305600407

McPherson, J. M., Popielarz, P. A., & Drobnic, S. (1992). Social networks and organizational dynamics. *American Sociological Review*, 153–170.

Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication. *Journal of Medical Internet Research*, *15*(4), e85. https://doi.org/10.2196/jmir.1933

Morin, C., Mercier, A., & Atlani-Duault, L. (2019). Text–Image Relationships in Tweets: Shaping the Meanings of an Epidemic. *Societies*, *9*(1), 12. https://doi.org/10.3390/soc9010012

Murthy, D., Gross, A., & McGarry, M. (2016). Visual Social Media and Big Data. Interpreting Instagram Images Posted on Twitter. *Digital Culture & Society*, *2*(2), 113–134. https://doi.org/10.14361/dcs-2016-0208

Newman, M. E. J. (2001). Clustering and preferential attachment in growing networks. *Physical Review E*, *64*(2). https://doi.org/10.1103/PhysRevE.64.025102

Nikunen, K. (2018). From Irony to Solidarity: Affective Practice and Social Media Activism. *Studies of Transition States and Societies*, *10*(2), 10–21.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, *11*(122–129), 1–2.

Olteanu, A., Weber, I., & Gatica-Perez, D. (2015). Characterizing the Demographics Behind the #BlackLivesMatter Movement. *ArXiv:1512.05671 [Cs]*. http://arxiv.org/abs/1512.05671

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*(3), 315.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & others. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Perrin, A., & Anderson, M. (2019). Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/

Ray, R., Brown, M., Fraistat, N., & Summers, E. (2017). Ferguson and the death of Michael Brown on Twitter: #BlackLivesMatter, #TCOT, and the evolution of collective identities. *Ethnic and Racial Studies*, *40*(11), 1797–1813. https://doi.org/10.1080/01419870.2017.1335422

Raynauld, V., Richez, E., & Morris, K. B. (2018). Canada is #IdleNoMore: Exploring dynamics of Indigenous political and civic protest in the Twitterverse. *Information, Communication & Society*, *21*(4), 626–642. https://doi.org/10.1080/1369118X.2017.1301522

Reiman, J., & Leighton, P. (2015). *Rich Get Richer and the Poor Get Prison, The (Subscription): Ideology, Class, and Criminal Justice*. Routledge.

Rodriguez, P. A., Drenkow, N., DeMenthon, D., Koterba, Z., Kauffman, K., Cornish, D., Paulhamus, B., & Vogelstein, R. J. (2014). Selection of universal features for image classification. *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference On*, 355–362. http://ieeexplore.ieee.org/abstract/document/6836078/

Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. *Proceedings of the 20th International Conference on World Wide Web*, 695–704. https://doi.org/10.1145/1963405.1963503

Schulhofer, S. J., Tyler, T. R., & Huq, A. Z. (2011). American policing at a crossroads: Unsustainable policies and the procedural justice alternative. *The Journal of Criminal Law and Criminology*, 335–374.

Schwämmlein, E., & Wodzicki, K. (2012). What to Tell About Me? Self-Presentation in Online Communities. *Journal of Computer-Mediated Communication*, *17*(4), 387–407. https://doi.org/10.1111/j.1083-6101.2012.01582.x

Scott, K. (2015). The pragmatics of hashtags: Inference and conversational style on Twitter. *Journal of Pragmatics*, *81*, 8–20. https://doi.org/10.1016/j.pragma.2015.03.015

Shapiro, D. L., Brass, D., & Labianca, J. (2008). EXAMINING JUSTICE FROM A SOCIAL NETWORK PERSPECTIVE. *Justice, Morality, and Social Responsibility*, 201.

Siegel, D. A. (2009). Social Networks and Collective Action. *American Journal of Political Science*, *53*(1), 122–138. https://doi.org/10.1111/j.1540-5907.2008.00361.x

Snow, D. A., Benford, R. D., & others. (1988). Ideology, frame resonance, and participant mobilization. *International Social Movement Research*, *1*(1), 197–217.

Snow, D. A., Soule, S. A., & Kriesi, H. (2008). Mapping the Terrain. In *The Blackwell companion to social movements* (pp. 3–16). John Wiley & Sons. https://books.google.com/books?hl=en&lr=&id=VYrPtQGrKkIC&oi=fnd&pg=PR5&dq=%22D.+Morris+and+Suzanne%22+%22Transnational+Processes+and%22+%22Collective+Identity,+Solidarity,+and%22+%22Diffusion+Processes+within+and+across%22+%22A.+Hunt+and+Robert+D.%22+%22Networks+and%22+%22della+Porta+and+Olivier%22+&ots=nrMJznj0Jc&sig=UhvrVUZpVzqMcIL-MI-C3LvdCjY

Starbird, K., & Palen, L. (2011). Voluntweeters: Self-organizing by digital volunteers in times of crisis. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, 1071–1080.

Stekelenburg, J. van. (2012). The occupy movement: Product of this time. *Development*, *55*(2), 224–231. https://doi.org/10.1057/dev.2012.6

Stekelenburg, J. van, Roggeband, C., & Klandermans, B. (2013). *The Future of Social Movement Research: Dynamics, Mechanisms, and Processes*. U of Minnesota Press.

Stewart, C. J., Denton, R. E. D., & Smith, C. A. (2012a). Languaging Strategies and Tactics of Social Movements. In *Persuasion and Social Movements, Sixth Edition* (6 edition, pp. 155–180). Waveland Pr Inc.

Stewart, C. J., Denton, R. E. D., & Smith, C. A. (2012b). Resisting Social Movements. In *Persuasion and Social Movements, Sixth Edition* (6 edition, pp. 321–342). Waveland Pr Inc.

Stewart, C. J., Denton, R. E. D., & Smith, C. A. (2012c). The Essential Characteristics of a Social Movement. In *Persuasion and Social Movements, Sixth Edition* (6 edition, pp. 1–26). Waveland Pr Inc.

Stewart, C. J., Denton, R. E. D., & Smith, C. A. (2012d). The Stages of Social Movements. In *Persuasion and Social Movements, Sixth Edition* (6 edition, pp. 85–110). Waveland Pr Inc.

Sunshine, J., & Tyler, T. R. (2003). The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing. *Law & Society Review*, *37*(3), 513–548. https://doi.org/10.1111/1540-5893.3703002

Sutton, J., Palen, L., & Shklovski, I. (2008). Backchannels on the front lines: Emergent uses of social media in the 2007 southern California wildfires. *Proceedings of the 5th International ISCRAM Conference*, 624–632. http://www.cs.colorado.edu/~palen/Papers/iscram08/BackchannelsISCRAM08.pdf

Tausczik, Y. R., & Pennebaker, J. W. (2009). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Taylor, V., Whittier, N., & others. (1995). Analytical approaches to social movement culture: The culture of the women's movement. *Social Movements and Culture*, *4*, 163–187.

Theocharis, Y. (2015). The Conceptualization of Digitally Networked Participation. *Social Media + Society*, *1*(2), 2056305115610140. https://doi.org/10.1177/2056305115610140

Thorne, J., Chen, M., Myrianthous, G., Pu, J., Wang, X., & Vlachos, A. (2017). Fake news stance detection using stacked ensemble of classifiers. *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing Meets Journalism*, 80–83.

Tonry, M. (2010). The Social, Psychological, and Political Causes of Racial Disparities in the American Criminal Justice System. *Crime and Justice*, *39*(1), 273–312. https://doi.org/10.1086/653045

Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *ResearchGate*. https://www.researchgate.net/publication/261217624_Big_Questions_for_Social_Media_Big_Data_Representativeness_Validity_and_Other_Methodological_Pitfalls

Turner, J. C., Brown, R. J., & Tajfel, H. (1979). Social comparison and group interest in ingroup favouritism. *European Journal of Social Psychology*, *9*(2), 187–204. https://doi.org/10.1002/ejsp.2420090207

Tyler, T. R. (1994). Psychological models of the justice motive: Antecedents of distributive and procedural justice. *Journal of Personality and Social Psychology*, *67*(5), 850–863. https://doi.org/10.1037/0022-3514.67.5.850

Valente, T. W. (1996). Social network thresholds in the diffusion of innovations. *Social Networks*, *18*(1), 69–89.

Valenzuela, S., Arriagada, A., & Scherman, A. (2014). Facebook, Twitter, and youth engagement: A quasi-experimental study of social media use and protest behavior using propensity score matching. *International Journal of Communication*, *8*, 25.

Van Laer, J., & Van Aelst, P. (2009). Cyber-protest and civil society: The Internet and action repertoires in social movements. *Handbook on Internet Crime*, 230–254.

van Stekelenburg, J. (2013). The Political Psychology of Protest. *European Psychologist*, *18*(4), 224–234. https://doi.org/10.1027/1016-9040/a000156

Varol, O., Ferrara, E., Ogan, C. L., Menczer, F., & Flammini, A. (2014). Evolution of Online User Behavior During a Social Upheaval. *Proceedings of the 2014 ACM Conference on Web Science*, 81–90. https://doi.org/10.1145/2615569.2615699

Vie, S. (2014). In defense of "slacktivism": The Human Rights Campaign Facebook logo as digital activism. *First Monday*, *19*(4). http://pear.accc.uic.edu/ojs/index.php/fm/article/view/4961

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). *Microblogging during two natural hazards events: What twitter may contribute to situational awareness*. *2*, 1079–1088. https://doi.org/10.1145/1753326.1753486

Vieweg, S., Palen, L., Liu, S. B., Hughes, A. L., & Sutton, J. (2008). Collective intelligence in disaster: An examination of the phenomenon in the aftermath of the 2007 Virginia Tech shootings. *Proceedings of the Information Systems for Crisis Response and Management Conference (ISCRAM)*. http://works.bepress.com/cgi/viewcontent.cgi?article=1017&context=vieweg

Vissers, S., & Stolle, D. (2014). Spill-Over Effects Between Facebook and On/Offline Political Participation? Evidence from a Two-Wave Panel Study. *Journal of Information Technology & Politics*, *11*(3), 259–275. https://doi.org/10.1080/19331681.2014.888383

Vitak, J., Zube, P., Smock, A., Carr, C. T., Ellison, N., & Lampe, C. (2010). It's Complicated: Facebook Users' Political Participation in the 2008 Election. *Cyberpsychology, Behavior, and Social Networking*, *14*(3), 107–114. https://doi.org/10.1089/cyber.2009.0226

Vitak, J., Zube, P., Smock, A., Carr, C. T., Ellison, N., & Lampe, C. (2011). It's complicated: Facebook users' political participation in the 2008 election. *Cyberpsychology, Behavior and Social Networking*, *14*(3), 107–114. https://doi.org/10.1089/cyber.2009.0226

Watts, D. J. (2007). A twenty-first century science. *Nature*, *445*(7127), 489. https://doi.org/10.1038/445489a

Welles, B. F., & Jackson, S. J. (2019). The Battle for #Baltimore: Networked Counterpublics and the Contested Framing of Urban Unrest. *International Journal of Communication*, *13*(0), 21.

Wellman, B., Haase, A. Q., Witte, J., & Hampton, K. (2001). Does the Internet Increase, Decrease, or Supplement Social Capital? *American Behavioral Scientist*, *45*(3), 436–455. https://doi.org/10.1177/00027640121957286

Wesslen, R., Nandu, S., Eltayeby, O., Gallicano, T., Levens, S., Jiang, M., & Shaikh, S. (2018). Bumper Stickers on the Twitter Highway: Analyzing the Speed and Substance of Profile Changes. *Proceedings of the International AAAI Conference on Web and Social Media*, *12*(1), Article 1. https://ojs.aaai.org/index.php/ICWSM/article/view/15056

Wilson, K. J., & Cohen, E. L. (2019). Do good causes leave bad impressions? Exploring the impact of photo frames and newsfeed updates on social impressions of facebook users. *Communication Research Reports*, *36*(4), 359–369. https://doi.org/10.1080/08824096.2019.1670632

Yates, D., & Paquette, S. (2011). Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *International Journal of Information Management*, *31*(1), 6–13. https://doi.org/10.1016/j.ijinfomgt.2010.10.001

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning Deep Features for Scene Recognition using Places Database. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 27* (pp. 487–495). Curran Associates, Inc. http://papers.nips.cc/paper/5349-learning-deep-features-for-scene-recognition-using-places-database.pdf

Zuckerman, E. (2014). New Media, New Civics? *Policy & Internet*, *6*(2), 151–168. https://doi.org/10.1002/1944-2866.POI360