# Online Appendix

## Appendix A: Historical PowerTrack API

As part of the Twitter data collection process, we utilized the Historical PowerTrack (HPT) API. This API provides access to the full historical archive of public Twitter data based on designated rules and operators. Therefore, we extracted the entire archive of tweets posted by the CRC and the CRC-A. This API also provides metadata about each tweet (e.g., timestamp, text) along with metadata about the account that posted the tweet. The account metadata include profile statistics (e.g., follower count) captured at the time that each tweet was posted. We also used the HPT API to collect retweets of the CRC's and the CRC-A's tweets, mentions of both accounts, and tweets from the general public with keywords and hashtags related to the Fort McMurray wildfire. The table below lists the full set of rules and operators we defined to extract data from the HPT API.

**Table A1 HPT API Rules and Operators**

| Rule Type | Rule |
|---|---|
| Date Range | 4/30/2016 - 8/4/2016 |
| From | from:redcrosscanada OR from:RedCrossAB |
| Retweets of | retweets_of:redcrosscanada OR retweets_of:RedCrossAB |
| Replies to | to:redcrosscanada OR to:RedCrossAB |
| Mentions of | @redcrosscanada OR @RedCrossAB |
| Hashtags | #ymmfire OR #yymfire OR #ABfire OR #ABfires OR #FortMcMurray OR #YMM OR #FortMac OR #ymmstrong OR #abstrong OR #albertastrong OR #albertafire OR #albertafires OR #fortmacfire OR #ymmhelp |
| Keywords | "Fort McMurray" OR "Fort Mac" OR "Red Cross" OR RedCross OR "Alberta Fire" OR "AB Fire" OR wildfire OR evacuee |

## Appendix B: Topic Modeling

To characterize a tweet's audience as disaster victims or supporters, we first determine the topics underlying the tweets in our sample. We rely on the topic modeling technique of Gibbs sampling algorithm for Dirichlet Multinomial Mixture (GSDMM) model proposed by (Yin and Wang 2014). Unlike traditional topic modeling algorithms (e.g., latent Dirichlet allocation (LDA) by Blei et al. (2003)) that view each document as a mixture or distribution of topics, GSDMM assumes that each document contains one topic. Naturally, this assumption is more fitting for short text data, such as tweets (Qiang et al. 2020). For this reason, our main results are derived from a GSDMM model, but we provide results based on an LDA model as a robustness check (see Appendix C). We find that our results are largely unaffected by the choice of topic modeling approach.

To evaluate topic models, we need objective measures for the quality of the results. Topic coherence, which measures the degree of semantic similarity among the top features in the topic, is often used for this purpose. We use two different types of topic coherence measures: the intrinsic *UMass* measure (Mimno et al. 2011) and the extrinsic *UCI* measure (Newman et al. 2010). The *UMass* measure uses as the pairwise score function:

$$\log \frac{D(w_i, w_j) + 1}{D(w_j)}$$

where $D(w_i, w_j)$ is the co-document frequency of words $w_i$ and $w_j$ and $D(w_j)$ is the document frequency of word $w_j$. The *UCI* measure treats an external corpus (e.g., Wikipedia) as a single meta-document and scores word pairs using pointwise mutual information:

$$\log \frac{p(w_i, w_j)}{p(w_i)p(w_j)}$$

where $p(w)$ is the probability of seeing word $w$ in a document and $p(w_i, w_j)$ is the probability of seeing both $w_i$ and $w_j$ in a document. These probabilities are empirically estimated using an external dataset.

### B.1. GSDMM

We estimate a GSDMM model for each number of topics between two and twenty inclusive. For each of these models, we compute the two coherence scores (i.e., *UMass* and *UCI* measures), normalize each score to a number between 0 and 1, and calculate their weighted average. The scores for the top five candidate number of topics are reported in Table A2 (sorted by the weighted scores in descending order).

**Table A2 Summary of Coherence Scores from GSDMM**

| Number of Topics ($n$) | Normalized UCI Scores | Normalized UMass Scores | Weighted Scores |
|---|---|---|---|
| 3 | 1 | 1 | 1 |
| 4 | 0.916 | 0.903 | 0.909 |
| 5 | 0.788 | 0.801 | 0.795 |
| 6 | 0.565 | 0.572 | 0.569 |
| 7 | 0.422 | 0.449 | 0.435 |

Based on the coherence scores, we determine that the best number of topics for our dataset is three. Table A3 lists some representative features for each of these three topics. The first topic includes words that are related to updates on the disaster, and the second topic concerns appeals for donations. Lastly, the third topic has features often used in instructions for registering to receive aid.

**Table A3 Key Features of Topics Identified from GSDMM**

| Topic 1 | Topic 2 | Topic 3 |
|---|---|---|
| support | donate | please |
| help | support | call |
| #ymmfire | text | register |
| #abfire | relief | email |
| #redcross | fundraising | contact |
| volunteer | help | patience |
| #ymm | appeal | soon |
| share | cost | information |
| evacuee | recovery | office |
| mcmurray | fund | confirm |

### B.2. LDA

As a robustness check, we run an LDA model to identify the audience of tweets. Table A4 summarizes the coherence scores from LDA models for each number of topics between two and twenty. Using the coherence scores, we determine the optimal number of topics to equal three. Table A5 provides some representative features for each of the three topics from our LDA model. The first topic concerns updates on the disaster and the recovery process, and the second topic contains words related to donation appeals. The third topic relates to registration for relief funds and aid. Therefore, the topics and key features resulting from an LDA model are comparable to those from a GSDMM model (displayed in A3). In Appendix C, we show that our main empirical results using a GSDMM model are robust to the alternative topic modeling results derived from an LDA model.

### B.3. Comparison between GSDMM and LDA

We next compare coherence scores from GSDMM and LDA models to validate our choice of a GSDMM model for the main analysis. Table A6 summarizes the coherence scores from the best-performing GSDMM and LDA models on our Twitter dataset from the 2016 Fort McMurray wildfire. We find that the topic model with the highest coherence score is the GSDMM model with three topics, which is the one used in our main analysis. In general, we also observe that the performance of GSDMM models is better than that of LDA models. The performance of the LDA model with three topics, however, is comparable to that of the best GSDMM models. This is why we use the LDA model with three topics for the robustness check in which we test the sensitivity of our results to an alternate topic modeling approach.

**Table A4 Summary of Coherence Scores from LDA**

| Number of Topics ($n$) | Normalized UCI Scores | Normalized UMass Scores | Weighted Scores |
|---|---|---|---|
| 3 | 1 | 1 | 1 |
| 4 | 0.723 | 0.747 | 0.735 |
| 5 | 0.670 | 0.710 | 0.690 |
| 6 | 0.627 | 0.669 | 0.648 |
| 7 | 0.533 | 0.566 | 0.550 |

**Table A5 Key Features of Topics Identified from LDA**

| Topic 1 | Topic 2 | Topic 3 |
|---|---|---|
| #redcross | donate | please |
| #abfire | support | register |
| support | help | call |
| help | fire | email |
| #ymmfire | affect | information |
| share | text | contact |
| need | fund | number |
| volunteer | relief | reach |
| emergency | online | message |
| assist | appeal | patience |

**Table A6 Comparing Coherence Scores from GSDMM and LDA**

| Model | Number of Topics ($n$) | Normalized UCI Scores | Normalized UMass Scores | Weighted Scores |
|---|---|---|---|---|
| GSDMM | 3 | 1 | 1 | 1 |
| GSDMM | 4 | 0.976 | 0.974 | 0.975 |
| GSDMM | 5 | 0.940 | 0.946 | 0.943 |
| LDA | 3 | 0.941 | 0.914 | 0.927 |
| GSDMM | 6 | 0.877 | 0.884 | 0.880 |
| GSDMM | 7 | 0.836 | 0.851 | 0.843 |
| GSDMM | 9 | 0.784 | 0.781 | 0.782 |
| GSDMM | 8 | 0.782 | 0.775 | 0.778 |
| GSDMM | 10 | 0.716 | 0.729 | 0.723 |
| LDA | 4 | 0.680 | 0.683 | 0.682 |

### B.4. Validation of Audience

The three topics identified by the GSDMM model are labeled as: updates on the disaster and recovery process, donation appeals, and instructions for registering for and receiving aid. As explained in Section 3.3, we set the audience of tweets belonging to the first two topics as supporters and to the third topic as disaster victims. We check the validity of tweets' audience by having two independent raters assess the audiences for 100 tweets that are randomly selected from our sample. The raters' instructions are to determine the audience of each tweet as disaster victims or supporters.

The inter-rater agreement, measured as the percentage of tweets for which both raters choose the same audience, is equal to 98%. In addition, the agreement between tweets' audience as determined by the GSDMM model and the raters is also high with an average agreement of 92%. More specifically, the agreement between the GSDMM model and the first rater is 93% and between the GSDMM model and the second rater is 91%. We, therefore, conclude that our process of determining the audience of tweets as disaster victims or supporters is valid.

## Appendix C: Robustness Checks

The results of the robustness checks are provided below. The first set of robustness checks relates to our topic modeling approach (Appendix C.1). We identify the audience of tweets using a latent Dirichlet allocation model (M6 in Tables A7, A8, and A9) and allocate the tweets belonging to the topic of updates about the disaster to victims instead of supporters (M7 in Tables A7, A8, and A9). The results of these analyses are consistent with the main results.

The second set of robustness checks pertains to alternative specifications to our models (Appendix C.1). The results of these are displayed in M8 through M12 in Tables A7, A8, and A9. First, we incorporate additional lags of the main variables of interest through $t-3$ (M8). Second, we use an alternate cutoff date to differentiate between the response and recovery phases (M9). Third, we add a variable for wind speed in the affected region as a supplemental proxy for the amount of interest in the wildfire (M10). Fourth, we verify that our results are robust to the inclusion of tweets that are not related to the focal disaster by dropping 190 tweets that are not disaster-related (M11). Fifth, we re-estimate the content creation decisions model using a logit specification and the engagement model using a Poisson specification (M12). Reassuringly, all of the results of these robustness checks are qualitatively consistent with the original results.

The third set of robustness checks is for the engagement model only (Appendix C.2). We make sure that our results are robust to engagement being defined as each of the individual engagement metrics included in the measurement of $Engagements_{i,t}$. Specifically, we re-estimate the engagement model with the dependent variable as the number of retweets, likes, replies, link clicks, media clicks, and hashtag clicks (M13 through M18 in Table A10). Moreover, to account for potential autocorrelation, we incorporate the lagged value for $Engagements_{i,t}$ (M19 in Table A11) and use robust clustered standard errors at the account level (M20 in Table A11). The results of these estimations are highly similar to the original results. Finally, we check that our results are not driven by the cases in which $AudienceMatch_{i,t}$ equals 1 but the accounts did not post (i.e., audience of $i$ in $t$ and $j$ in $t-1$ is *None*). We include a binary variable ($BothTweeted_{i,t}$) set to 1 when $i$ posted in $t$ and $j$ posted in $t-1$ and 0 otherwise. This allows us to separate the cases in which the targeted audience of both accounts was *None* vs. *Victims*, *Supporters*, or *Both*. The results of this robustness check (M21 in Table A11) indicate that $BothTweeted_{i,t}$ is positive but not significant (0.096, $p > 0.10$). Thus, we do not find evidence that engagement is different when both accounts target *None* (i.e., when both do not post). All other results remain consistent.

We also assemble a second test bed to show that our results are not sensitive to the focal disaster event in our study. Details about the dataset and analysis results are reported in Appendix C.3. We find that the results for the second test bed are qualitatively consistent with our original set of results.

## C.1. Alternative Topic Modeling Approach and Specifications

**Table A7 Robustness Checks for the Whether to Post Model - Alternative Topic Modeling and Specifications**

| | M6 LDA | M7 Updates to Victims | M8 Add Lags | M9 Alternate Cutoff | M10 Wind Speed | M11 Disaster Tweets | M12 Logit |
|---|---|---|---|---|---|---|---|
| $Post_{i,t-1}$ | 0.559*** | 0.374*** | 0.449*** | 0.510*** | 0.599*** | 0.558*** | 0.959*** |
| | (0.068) | (0.068) | (0.069) | (0.068) | (0.067) | (0.074) | (0.129) |
| $Post_{j,t-1}$ | 0.306*** | 0.257*** | 0.336*** | 0.208** | 0.208** | 0.231** | 0.413** |
| | (0.089) | (0.090) | (0.089) | (0.093) | (0.094) | (0.102) | (0.189) |
| $National_i$ | 0.163*** | 0.324*** | 0.114* | 0.502*** | 0.474*** | 0.286** | 0.840*** |
| | (0.062) | (0.064) | (0.064) | (0.063) | (0.063) | (0.067) | (0.136) |
| $Post_{j,t-1}$ | -0.083 | -0.214 | -0.260* | -0.189 | -0.067 | -0.132 | -0.496* |
| $\times National_i$ | (0.143) | (0.144) | (0.144) | (0.144) | (0.144) | (0.158) | (0.288) |
| $Response_t$ | -0.116 | -0.158 | -0.300*** | -0.738*** | -0.170* | -0.031 | -0.323* |
| | (0.095) | (0.096) | (0.095) | (0.091) | (0.096) | (0.099) | (0.185) |
| $Mentions_{i,t}$ | 0.004* | 0.003 | 0.007*** | 0.002 | -2E-04 | 0.004* | -0.001 |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.004) |
| $TwitterInterest_t$ | -0.388*** | -0.326*** | -0.371*** | -0.272*** | -0.299*** | -0.327*** | 0.053* |
| | (0.017) | (0.018) | (0.018) | (0.017) | (0.018) | (0.018) | (0.028) |
| $GoogleTrends_t$ | 0.039*** | 0.042*** | 0.030*** | 0.041*** | 0.041*** | 0.029*** | 0.039*** |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.008) |
| $HrsSinceFire_t$ | -0.001*** | -0.001*** | -0.001*** | -0.001*** | -0.001*** | -0.001*** | -0.001*** |
| | (5E-05) | (5E-05) | (5E-05) | (6E-05) | (5E-05) | (6E-05) | (1E-04) |
| $Post_{i,t-2}$ | | | 0.049 | | | | |
| | | | (0.074) | | | | |
| $Post_{j,t-2}$ | | | 0.068 | | | | |
| | | | (0.076) | | | | |
| $Post_{i,t-3}$ | | | 0.114 | | | | |
| | | | (0.074) | | | | |
| $Post_{j,t-3}$ | | | 0.143* | | | | |
| | | | (0.076) | | | | |
| $WindSpeed_t$ | | | | | 0.003 | | |
| | | | | | (0.004) | | |
| Intercept | 1.554*** | 1.139*** | 1.661*** | 1.251*** | 0.788*** | 1.296*** | -1.327*** |
| | (0.086) | (0.101) | (0.088) | (0.076) | (0.078) | (0.083) | (0.032) |

*Note:* Standard errors are presented in parentheses. Hour fixed effects are included but not shown. The number of observations is 4,546. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

**Table A8 Robustness Checks for the Who to Post for Model - Alternative Topic Modeling and Specifications**

| | M6 LDA | M7 Updates to Victims | M8 Add Lags | M9 Alternate Cutoff | M10 Wind Speed | M11 Disaster Tweets | M12 Logit |
|---|---|---|---|---|---|---|---|
| $Victim_{i,t-1}$ | 0.343*** | 0.147 | 0.368*** | 0.354*** | 0.249** | 0.209 | 0.372* |
| | (0.124) | (0.142) | (0.129) | (0.124) | (0.124) | (0.132) | (0.225) |
| $Victim_{j,t-1}$ | 0.229** | 0.077 | 0.259*** | 0.161* | 0.188* | 0.364*** | 0.810*** |
| | (0.102) | (0.108) | (0.099) | (0.097) | (0.096) | (0.104) | (0.180) |
| $Support_{i,t-1}$ | -0.068 | -0.133 | 0.033 | -0.028 | 0.002 | 0.240* | 0.559** |
| | (0.118) | (0.132) | (0.124) | (0.121) | (0.120) | (0.138) | (0.236) |
| $Support_{j,t-1}$ | 0.091 | -0.197*** | 0.043 | -0.027 | 3E-04 | 0.032 | 0.056 |
| | (0.105) | (0.073) | (0.109) | (0.111) | (0.108) | (0.132) | (0.226) |
| $National_i$ | -0.364*** | -0.527*** | -0.524*** | -0.465*** | -0.413*** | -0.767*** | -1.577*** |
| | (0.108) | (0.138) | (0.118) | (0.114) | (0.112) | (0.129) | (0.231) |
| $Victim_{j,t-1}$ | -0.118* | -0.114 | -0.207*** | -0.231*** | -0.193*** | -0.342*** | -0.813*** |
| $\times National_i$ | (0.071) | (0.091) | (0.079) | (0.073) | (0.071) | (0.078) | (0.132) |
| $Support_{j,t-1}$ | -0.109 | -0.116** | -0.041 | -0.087 | -0.032 | -0.006 | 0.026 |
| $\times National_i$ | (0.067) | (0.050) | (0.074) | (0.074) | (0.069) | (0.085) | (0.149) |
| $Response_t$ | -0.492*** | -0.435*** | -0.748*** | -0.261* | -0.571*** | -0.526*** | -0.605* |
| | (0.124) | (0.057) | (0.069) | (0.135) | (0.118) | (0.169) | (0.348) |
| $Mentions_{i,t}$ | -0.004* | 0.003 | 0.002 | 8E-05 | -0.002 | -0.001 | 0.005 |
| | (0.002) | (0.004) | (0.003) | (0.002) | (0.002) | (0.003) | (0.004) |
| $TwitterInterest_t$ | 0.126*** | 0.381*** | 0.172*** | 0.146*** | 0.123*** | 0.095*** | -0.537*** |
| | (0.024) | (0.029) | (0.023) | (0.028) | (0.028) | (0.030) | (0.055) |
| $GoogleTrends_t$ | -0.005 | -0.018*** | -0.008 | -0.016*** | -0.005 | -0.013** | -0.011 |
| | (0.005) | (0.007) | (0.005) | (0.005) | (0.005) | (0.005) | (0.009) |
| $HrsSinceFire_t$ | -5E-04*** | 8E-05 | -8E-04*** | -7E-04*** | -7E-04*** | -4E-04*** | -0.002*** |
| | (9E-05) | (1E-04) | (9E-05) | (1E-04) | (1E-04) | (1E-04) | (2E-04) |
| $Victim_{i,t-2}$ | | | 0.065 | | | | |
| | | | (0.135) | | | | |
| $Victim_{j,t-2}$ | | | -0.049 | | | | |
| | | | (0.147) | | | | |
| $Support_{i,t-2}$ | | | -0.324** | | | | |
| | | | (0.134) | | | | |
| $Support_{j,t-2}$ | | | -0.184 | | | | |
| | | | (0.148) | | | | |
| $Victim_{i,t-3}$ | | | 0.325** | | | | |
| | | | (0.144) | | | | |
| $Victim_{j,t-3}$ | | | 0.166 | | | | |
| | | | (0.133) | | | | |
| $Support_{i,t-3}$ | | | -0.383*** | | | | |
| | | | (0.134) | | | | |
| $Support_{j,t-3}$ | | | 0.015 | | | | |
| | | | (0.146) | | | | |
| $WindSpeed_t$ | | | | | 0.004 | | |
| | | | | | (0.007) | | |
| Intercept | 0.089*** | -0.497*** | 0.275*** | 0.229*** | 0.244*** | 0.632*** | 6.277*** |
| | (0.009) | (0.012) | (0.012) | (0.013) | (0.011) | (0.013) | (0.029) |

*Note:* Standard errors are presented in parentheses. Hour fixed effects are included but not shown. The number of observations is 4,546. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

**Table A9 Robustness Checks for the Engagement Model - Alternative Topic Modeling and Specifications**

| | M6 LDA | M7 Updates to Victims | M8 Add Lags | M9 Alternate Response | M10 Wind Speed | M11 Disaster Tweets | M12 Poisson |
|---|---|---|---|---|---|---|---|
| $AudienceMatch_{i,t}$ | -0.339*** | -0.304*** | -0.295*** | -0.390*** | -0.328*** | -0.523*** | -0.528*** |
| | (0.095) | (0.094) | (0.102) | (0.101) | (0.094) | (0.080) | (0.097) |
| $Response_t$ | 0.321** | 0.343*** | 0.306** | 0.132 | 0.241* | -0.058 | 0.019 |
| | (0.125) | (0.131) | (0.124) | (0.117) | (0.123) | (0.126) | (0.144) |
| $AudienceMatch_{i,t}$ $\times Response_t$ | 0.337*** | 0.306** | 0.365*** | 0.400*** | 0.375*** | 0.382*** | 0.405*** |
| | (0.127) | (0.134) | (0.124) | (0.126) | (0.127) | (0.122) | (0.110) |
| $VictimTweets_{i,t}$ | 0.002 | 0.019 | -0.002 | 0.005 | 0.021 | 0.041 | 0.007 |
| | (0.032) | (0.025) | (0.029) | (0.028) | (0.028) | (0.026) | (0.034) |
| $SupportTweets_{i,t}$ | -0.027 | -0.153** | -0.032 | -0.027 | -0.016 | -0.033 | 0.026 |
| | (0.034) | (0.061) | (0.038) | (0.039) | (0.040) | (0.044) | (0.029) |
| $VictimTweets_{j,t-1}$ | -0.083** | -0.074** | -0.105*** | -0.093*** | -0.099*** | -0.103*** | -0.063*** |
| | (0.035) | (0.031) | (0.027) | (0.027) | (0.027) | (0.028) | (0.021) |
| $SupportTweets_{j,t-1}$ | -0.088* | -0.108** | -0.046 | -0.043 | -0.017 | 0.027 | -0.037 |
| | (0.050) | (0.052) | (0.055) | (0.057) | (0.054) | (0.051) | (0.039) |
| $Interactive_{i,t}$ | 0.250*** | 0.271*** | 0.239*** | 0.251*** | 0.245*** | 0.186*** | 0.013 |
| | (0.047) | (0.052) | (0.049) | (0.050) | (0.052) | (0.052) | (0.039) |
| $Interactive_{j,t-1}$ | 0.015 | 0.008 | -0.017 | -0.018 | -0.048 | -0.029 | 0.038 |
| | (0.054) | (0.042) | (0.055) | (0.058) | (0.055) | (0.059) | (0.043) |
| $FollChange_{i,t}$ | 0.006*** | 0.006*** | 0.005** | 0.004** | 0.006** | 0.004* | 0.004*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.001) |
| $FollChange_{j,t}$ | 0.012*** | 0.012*** | 0.013*** | 0.012*** | 0.013*** | 0.007*** | 0.002* |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.001) |
| $Mentions_{i,t}$ | -0.005* | -0.005* | -0.004 | -0.003 | -0.006** | -0.010*** | 0.001 |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.002) |
| $Mentions_{j,t}$ | -0.004 | -0.005 | -0.004 | -0.004 | -0.006* | 0.007** | 0.010*** |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.004) | (0.003) | (0.002) |
| $TwitterInterest_t$ | 0.550*** | 0.546*** | 0.546*** | 0.584*** | 0.547*** | 0.701*** | 0.583*** |
| | (0.044) | (0.044) | (0.044) | (0.044) | (0.043) | (0.041) | (0.046) |
| $GoogleTrends_t$ | 0.019*** | 0.020*** | 0.019*** | 0.021*** | 0.024*** | 0.017*** | -0.003 |
| | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.004) | (0.006) |
| $HrsSinceFire_t$ | -3E-05 | -4E-05 | -4E-05 | 3E-05 | -9E-05 | -0.001*** | -2E-04* |
| | (9E-05) | (9E-05) | (9E-05) | (8E-05) | (8E-05) | (8E-05) | (1E-04) |
| First lag† for $AudienceMatch_{i,t}$ | | | -0.016 | | | | |
| | | | (0.091) | | | | |
| Second lag‡ for $AudienceMatch_{i,t}$ | | | -0.085 | | | | |
| | | | (0.088) | | | | |
| $WindSpeed_t$ | | | | | -0.018*** | | |
| | | | | | (0.004) | | |
| Intercept | -2.733*** | -2.738*** | -2.667*** | -2.947*** | -2.421*** | -3.929*** | -2.145*** |
| | (0.336) | (0.335) | (0.340) | (0.334) | (0.324) | (0.308) | (0.341) |

*Notes:* Robust standard errors are presented in parentheses. Account fixed effects and hour fixed effects are included but not shown. The number of observations is 4,546. †The first lag for $AudienceMatch_{i,t}$ evaluates whether the audience for $i$ in $t$ matches the audience for $j$ in $t-2$. ‡The second lag for $AudienceMatch_{i,t}$ evaluates whether the audience for $i$ in $t$ matches the audience for $j$ in $t-3$. $^{*}p<0.10$; $^{**}p<0.05$; $^{***}p<0.01$

## C.2. Robustness Checks for the Engagement Model

**Table A10 Robustness Checks for the Engagement Model - DV as Individual Engagement Metrics**

| | M13 DV: Retweets | M14 DV: Likes | M15 DV: Replies | M16 DV: Link Clicks | M17 DV: Media Clicks | M18 DV: Hashtag Clicks |
|---|---|---|---|---|---|---|
| $AudienceMatch_{i,t}$ | -0.761*** | -0.293*** | -0.918*** | -0.419*** | -0.245* | -0.300*** |
| | (0.073) | (0.073) | (0.155) | (0.082) | (0.128) | (0.091) |
| $Response_t$ | 0.164 | 0.425*** | -0.329* | 0.134 | -0.381*** | -0.030 |
| | (0.140) | (0.142) | (0.177) | (0.130) | (0.134) | (0.134) |
| $AudienceMatch_{i,t}$ | 0.845*** | 0.229* | 0.657*** | 0.427*** | 0.269* | 0.458*** |
| $\times Response_t$ | (0.130) | (0.131) | (0.181) | (0.117) | (0.161) | (0.138) |
| $VictimTweets_{i,t}$ | -0.042 | -0.028 | 0.368*** | 0.016 | -0.204*** | -0.008 |
| | (0.030) | (0.028) | (0.030) | (0.035) | (0.058) | (0.034) |
| $SupportTweets_{i,t}$ | -0.064 | -0.045 | 0.140*** | -0.077* | -0.038 | -0.117*** |
| | (0.042) | (0.040) | (0.044) | (0.045) | (0.058) | (0.037) |
| $VictimTweets_{j,t-1}$ | -0.075** | -0.087*** | -0.040 | -0.051** | -0.099*** | -0.078*** |
| | (0.035) | (0.032) | (0.029) | (0.025) | (0.033) | (0.026) |
| $SupportTweets_{j,t-1}$ | -0.066 | -0.052 | -0.022 | 0.067 | -0.177** | -0.127*** |
| | (0.055) | (0.067) | (0.044) | (0.049) | (0.072) | (0.048) |
| $Interactive_{i,t}$ | 0.263*** | 0.193*** | -0.099** | 0.233*** | 0.282*** | 0.147*** |
| | (0.051) | (0.047) | (0.049) | (0.066) | (0.080) | (0.046) |
| $Interactive_{j,t-1}$ | 0.001 | 0.067 | 0.027 | -0.119*** | 0.045 | 0.104 |
| | (0.063) | (0.071) | (0.053) | (0.042) | (0.073) | (0.068) |
| $FollChange_{i,t}$ | 0.011*** | 0.008*** | 0.001 | 0.006** | 0.014** | 0.009*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| $FollChange_{j,t}$ | 0.011*** | 0.008*** | 0.001 | 0.006** | 0.014** | 0.009*** |
| | (0.004) | (0.002) | (0.002) | (0.002) | (0.005) | (0.003) |
| $Mentions_{i,t}$ | -0.012*** | -0.005 | 0.007** | 3E-04 | 0.003 | -0.008*** |
| | (0.003) | (0.003) | (0.003) | (0.002) | (0.004) | (0.002) |
| $Mentions_{j,t}$ | 0.001 | 0.001 | 0.004 | 0.006* | -0.021*** | 0.001 |
| | (0.004) | (0.003) | (0.002) | (0.003) | (0.006) | (0.004) |
| $TwitterInterest_t$ | 0.606*** | 0.400*** | 0.695*** | 0.718*** | 0.547*** | 0.749*** |
| | (0.041) | (0.037) | (0.078) | (0.049) | (0.061) | (0.053) |
| $GoogleTrends_t$ | 0.030*** | 0.007 | -0.012** | 0.023*** | 0.022*** | 0.027*** |
| | (0.005) | (0.005) | (0.006) | (0.004) | (0.007) | (0.006) |
| $HrsSinceFire_t$ | -1E-04 | -0.001*** | -3E-04 | 2E-04** | 2E-05 | -1E-04 |
| | (7E-05) | (7E-05) | (2E-04) | (8E-05) | (1E-04) | (1E-04) |
| Intercept | -4.678*** | -2.661*** | -5.928*** | -6.023*** | -3.654*** | -6.978*** |
| | (0.314) | (0.278) | (0.558) | (0.342) | (0.480) | (0.382) |

*Notes:* DV, dependent variable. Robust standard errors are presented in parentheses. Account fixed effects and hour fixed effects are included but not shown. The number of observations is 4,546. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

**Table A11 Robustness Checks for the Engagement Model - Alternate Specifications**

| | M19<br>Add DV lag | M20<br>Clustered SEs | M21<br>Both Tweeted |
|---|---|---|---|
| $AudienceMatch_{i,t}$ | -0.352*** | -0.351*** | -0.360*** |
| | (0.089) | (0.108) | (0.097) |
| $Response_t$ | 0.188 | 0.292 | 0.289** |
| | (0.119) | (0.505) | (0.126) |
| $AudienceMatch_{i,t} \times Response_t$ | 0.414*** | 0.393*** | 0.388*** |
| | (0.120) | (0.012) | (0.127) |
| $VictimTweets_{i,t}$ | 0.011 | 0.003 | -0.003 |
| | (0.028) | (0.069) | (0.029) |
| $SupportTweets_{i,t}$ | -0.033 | -0.027 | -0.035 |
| | (0.040) | (0.101) | (0.040) |
| $VictimTweets_{j,t-1}$ | -0.078*** | -0.103*** | -0.111*** |
| | (0.029) | (0.003) | (0.031) |
| $SupportTweets_{j,t-1}$ | 0.004 | -0.042 | -0.050 |
| | (0.046) | (0.085) | (0.057) |
| $Interactive_{i,t}$ | 0.266*** | 0.248*** | 0.246*** |
| | (0.052) | (0.059) | (0.049) |
| $Interactive_{j,t-1}$ | -0.099* | -0.023 | -0.021 |
| | (0.055) | (0.159) | (0.057) |
| $FollChange_{i,t}$ | -0.009*** | 0.005*** | 0.006** |
| | (0.003) | (2E-04) | (0.002) |
| $FollChange_{j,t}$ | 0.012*** | 0.013*** | 0.013*** |
| | (0.003) | (0.002) | (0.003) |
| $Mentions_{i,t}$ | -0.003 | -0.004 | -0.004 |
| | (0.003) | (0.006) | (0.003) |
| $Mentions_{j,t}$ | -0.002 | -0.004 | -0.004 |
| | (0.003) | (0.007) | (0.004) |
| $TwitterInterest_t$ | 0.511*** | 0.548*** | 0.550*** |
| | (0.041) | (0.138) | (0.045) |
| $GoogleTrends_t$ | 0.019*** | 0.019 | 0.019*** |
| | (0.005) | (0.023) | (0.005) |
| $HrsSinceFire_t$ | -4E-05 | -4E-05 | -4E-05 |
| | (8E-05) | (0.001) | (9E-05) |
| $Engagements_{i,t-1}$ | 0.002*** | | |
| | (0.001) | | |
| $BothTweeted_{i,t}$ | | | 0.096 |
| | | | (0.116) |
| Intercept | -2.495*** | -2.716 | -2.718*** |
| | (0.310) | (2.059) | (0.337) |

*Notes:* DV, dependent variable. Robust standard errors are presented in parentheses. For M20, standard errors are clustered at the account level. Account fixed effects and hour fixed effects are included but not shown. The number of observations is 4,546. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

### C.3. Second Test Bed

We build a second test bed to confirm that our results are not derived from the disaster event featured in our study (i.e., the 2016 Fort McMurray wildfire). The disaster event for the second test bed is the 2017 British Columbia (BC) wildfires. We selected this disaster because its scope and impact were similar to the 2016 Fort McMurray wildfire. Since this disaster occurred in the BC province of Canada, the relevant national account represents the Canadian Red Cross (CRC) and the relevant local account represents the CRC in BC (i.e., CRC-BC).

The dataset for the second test bed is identical to the dataset used in our main analysis. That is, we collected Twitter data on all tweets published by the CRC and the CRC-BC as well as the engagement for these tweets. Like the original dataset, the data collection period for this dataset begins with the start of the disaster event and ends one month after the provincial state of emergency was lifted, thereby covering the response and recovery phases. Therefore, the second test bed dataset is fundamentally analogous to the original dataset (i.e., involves a national and a local account, spans a similar length of time, and contains the same variables).

Using the second test bed data, we estimate a GSDMM model for each number of topics between two and twenty inclusive. The scores for the top five candidate number of topics are reported in Table A12 (sorted by the weighted scores in descending order).

**Table A12 Summary of Coherence Scores from GSDMM**

| Number of Topics ($n$) | Normalized UCI Scores | Normalized UMass Scores | Weighted Scores |
|---|---|---|---|
| 3 | 1 | 1 | 1 |
| 7 | 0.366 | 0.248 | 0.307 |
| 8 | 0.390 | 0.186 | 0.288 |
| 5 | 0.279 | 0.215 | 0.247 |
| 9 | 0.308 | 0.179 | 0.243 |

Based on the coherence scores, we conclude that three is the best number of topics for our dataset. Table A13 lists representative features for each of these three topics. The topics and key features from a GSDMM model on the BC wildfire dataset are similar to those from a GSDMM model on the Fort McMurray wildfire dataset (Table A3). The first topic includes words related to updates on the disaster, and the second topic has features related to donation appeals. The third topic contains features concerning registration instructions. As before, we set the audience of tweets belonging to the first and second topics as supporters and to the third topic as disaster victims.

**Table A13 Key Features of Topics Identified from GSDMM**

| Topic 1 | Topic 2 | Topic 3 |
|---|---|---|
| help | thank | register |
| #bcwildfire | donate | call |
| people | support | please |
| support | help | contact |
| #bcfires | fund | message |
| volunteer | impact | sorry |
| evacuee | affect | evacuate |
| centre | appeal | dm |
| share | fundraising | information |
| #bcwildfires | relief | follow |

We then estimate the empirical models in our study. We note that when estimating the engagement model for the second test bed, we do not include $FollsChange_{i,t}$. The historical counts of followers were not available from Twitter at the time of data collection for the second test bed. The results from the second test bed are qualitatively consistent with the results from our main analysis. Below, Table A14 presents the results for the content creation decisions model and Table A15 reports the results for the engagement model.

**Table A14 Estimation Results for the Content Creation Decisions Model - Second Test Bed**

| | M22<br>DV: $Post_{i,t}$ | M23<br>DV: $Victim_{i,t}$ |
|---|---|---|
| $Post_{i,t-1}$ | 0.337*** | |
| | (0.082) | |
| $Post_{j,t-1}$ | 0.168* | |
| | (0.091) | |
| $Victim_{i,t-1}$ | | 0.127*** |
| | | (0.028) |
| $Victim_{j,t-1}$ | | 0.155*** |
| | | (0.011) |
| $Support_{i,t-1}$ | | -0.055 |
| | | (0.149) |
| $Support_{j,t-1}$ | | -0.006 |
| | | (0.039) |
| $National_i$ | 0.335*** | -0.148 |
| | (0.061) | (0.155) |
| $Post_{j,t-1} \times National_i$ | -0.185*** | |
| | (0.040) | |
| $Victim_{j,t-1} \times National_i$ | | -0.112*** |
| | | (0.005) |
| $Support_{j,t-1} \times National_i$ | | -0.033*** |
| | | (0.010) |
| $Response_t$ | -0.648*** | -0.033 |
| | (0.088) | (0.028) |
| $Mentions_{i,t}$ | 0.085*** | 0.048** |
| | (0.009) | (0.019) |
| $TwitterInterest_t$ | -0.401*** | -0.132*** |
| | (0.021) | (0.037) |
| $GoogleTrends_t$ | 0.012*** | 0.004 |
| | (0.002) | (0.004) |
| $HrsSinceFire_t$ | -0.001*** | -2E-04* |
| | (6E-05) | (1E-04) |
| Intercept | 0.946*** | 0.167*** |
| | (0.091) | (0.015) |

*Notes:* DV, dependent variable. Standard errors are presented in parentheses. Hour fixed effects are included but not shown. The number of observations is 4,820. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

**Table A15 Estimation Results for the Engagement Model - Second Test Bed**

| | M24 Overall Sample | M25 National Account Subsample | M26 Local Account Subsample |
|---|---|---|---|
| $AudienceMatch_{i,t}$ | -0.381$^{***}$ | -0.533$^{***}$ | -0.352 |
| | (0.136) | (0.092) | (0.246) |
| $Response_t$ | 1.290$^{***}$ | 0.344$^{***}$ | 2.780$^{***}$ |
| | (0.146) | (0.101) | (0.299) |
| $AudienceMatch_{i,t} \times Response_t$ | 0.413$^{**}$ | 0.408$^{***}$ | 0.698$^{**}$ |
| | (0.161) | (0.103) | (0.277) |
| $VictimTweets_{i,t}$ | 0.351$^{**}$ | 0.044 | 0.751$^{**}$ |
| | (0.151) | (0.076) | (0.351) |
| $SupportTweets_{i,t}$ | 0.331$^{**}$ | 0.127 | 1.464$^{***}$ |
| | (0.148) | (0.090) | (0.473) |
| $VictimTweets_{j,t-1}$ | 0.024 | 0.447$^{***}$ | -0.102 |
| | (0.131) | (0.121) | (0.188) |
| $SupportTweets_{j,t-1}$ | -0.144 | 0.415$^{**}$ | -0.281 |
| | (0.166) | (0.164) | (0.211) |
| $Interactive_{i,t}$ | 0.232 | 0.122 | 0.327 |
| | (0.151) | (0.087) | (0.395) |
| $Interactive_{j,t-1}$ | -0.123 | -0.506$^{***}$ | 0.128 |
| | (0.180) | (0.156) | (0.228) |
| $Mentions_{i,t}$ | 0.032$^{***}$ | 0.058$^{***}$ | 0.216$^{**}$ |
| | (0.009) | (0.008) | (0.093) |
| $Mentions_{j,t}$ | 0.091$^{***}$ | 0.058 | 0.050$^{***}$ |
| | (0.015) | (0.041) | (0.017) |
| $TwitterInterest_t$ | 0.231$^{***}$ | 0.151$^{***}$ | 0.310$^{***}$ |
| | (0.044) | (0.033) | (0.110) |
| $GoogleTrends_t$ | 0.006$^{***}$ | -3E-04 | 0.014$^{***}$ |
| | (0.002) | (0.001) | (0.004) |
| $HrsSinceFire_t$ | 3E-04$^{***}$ | 3E-05 | 0.001$^{***}$ |
| | (1E-04) | (8E-05) | (3E-04) |
| Intercept | -1.953$^{***}$ | 1.961$^{***}$ | -4.201$^{***}$ |
| | (0.307) | (0.252) | (0.830) |
| Account fixed effects | Yes | No | No |
| Hour fixed effects | Yes | Yes | Yes |
| Observations | 4,820 | 2,410 | 2,410 |

*Notes:* Robust standard errors are presented in parentheses. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

## References

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *Journal of machine Learning research* 3(Jan):993–1022.

Mimno D, Wallach H, Talley E, Leenders M, McCallum A (2011) Optimizing semantic coherence in topic models. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 262–272.

Newman D, Lau JH, Grieser K, Baldwin T (2010) Automatic evaluation of topic coherence. *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*, 100–108.

Qiang J, Qian Z, Li Y, Yuan Y, Wu X (2020) Short text topic modeling techniques, applications, and performance: a survey. *IEEE Transactions on Knowledge and Data Engineering* 34(3):1427–1445.

Yin J, Wang J (2014) A dirichlet multinomial mixture model-based approach for short text clustering. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 233–242.