

# HAWKES BINOMIAL TOPIC MODEL WITH APPLICATIONS TO COUPLED CONFLICT-TWITTER DATA

BY GEORGE MOHLER<sup>1</sup>, ERIN MCGRATH<sup>2</sup>, CODY BUNTAIN<sup>3</sup> AND GARY LAFREE<sup>4</sup>

<sup>1</sup>*Department of Computer and Information Science, Indiana University—Purdue University Indianapolis, [gmohler@iupui.edu](mailto:gmohler@iupui.edu)*

<sup>2</sup>*National Consortium for the Study of Terrorism and Responses to Terrorism, University of Maryland, [McGrath.erin@gmail.com](mailto:McGrath.erin@gmail.com)*

<sup>3</sup>*Department of Informatics, New Jersey Institute of Technology, [cbuntain@njit.edu](mailto:cbuntain@njit.edu)*

<sup>4</sup>*Department of Criminology and Criminal Justice, University of Maryland, [glafree@umd.edu](mailto:glafree@umd.edu)*

We consider the problem of modeling and clustering heterogeneous event data arising from coupled conflict event and social media data sets. In this setting conflict events trigger responses on social media, and, at the same time, signals of grievance detected in social media may serve as leading indicators for subsequent conflict events. For this purpose we introduce the Hawkes Binomial Topic Model (HBTM) where marks, Tweets and conflict event descriptions are represented as bags of words following a Binomial distribution. When viewed as a branching process, the daughter event bag of words is generated by randomly turning on/off parent words through independent Bernoulli random variables. We then use expectation–maximization to estimate the model parameters and branching structure of the process. The inferred branching structure is then used for topic cascade detection, short-term forecasting, and investigating the causal dependence of grievance on social media and conflict events in recent elections in Nigeria and Kenya.

**1. Background and motivation.** Twitter and other social media platforms have emerged as important tools for the public to communicate responses to crises and terrorist attacks (Buntain et al. (2016)) and, more generally, to communicate collectively, exchanging grievances that may catalyze mobilization (Bellin (2012), Rennick (2013)). Research has focused on understanding public sentiment around these types of events and determining the central actors in the social network that are key to shaping public response (Buntain et al. (2016)) along with measuring short-term changes in the intensity of conflict using social media (Zeitzoff (2011)) or considering effects from regional instability (Bhavnani and Donnay (2012)).

At a more macro spatial-temporal scale, recent research has focused on modeling the endogenous and exogenous processes that generate thousands of terrorist and conflict events at the level of countries and years or decades. For this purpose, point processes are used to model contagion effects in the risk of terrorist activity (Porter and White (2012)) and both contagion and exogenous rate fluctuations in conflict (Mohler (2013), Zammit-Mangion et al. (2012, 2013)). Because of the relative infrequency of conflict and terrorist events, having auxiliary data that can provide a signal for the risk of future events is highly desirable. While conflict events have been shown to influence overall regional instability (Bhavnani and Donnay (2012)), point process models of conflict to date have focused on univariate data (Porter and White (2012), Mohler (2013), Zammit-Mangion et al. (2012, 2013)). In this paper we use Twitter posts relevant to elections as auxiliary data to investigate whether we can provide a leading indicator for conflict activity and contentious events. Because only a subset of Tweets may be relevant to a particular event, we introduce a Hawkes topic model to capture Tweet content and conflict event descriptions.

---

Received September 2018; revised February 2020.

*Key words and phrases.* Hawkes process, dynamic topic model, social media, elections.

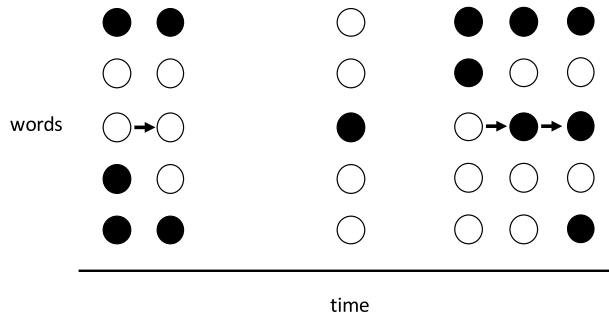


FIG. 1. In the HTBM, spontaneous events occur with marks generated by a binomial random variable over the dictionary of keywords contained in the data set. Events then trigger offspring events whose marks are generated by switching parent event words off with probability  $p_{\text{off}}$  and on with probability  $p_{\text{on}}$ .

We propose a Hawkes Binomial Topic Model (HBTM) where marks (Tweets and conflict event descriptions) are represented as bags of words following a Binomial distribution (see Figure 1). When viewed as a branching process, the daughter event bag of words is generated by randomly turning on/off parent words through independent Bernoulli random variables. We allow for a secondary mark, representing event type (Twitter or conflict event type) and extend the HBTM to a mutually exciting point process that captures crossexcitation effects between different types of conflict data along with social media data relevant to the intensity of conflict. We then analyze a unique merged data set of all conflict events from the Armed Conflict and Location Event Data Project (ACLED) (Raleigh et al. (2010)) from a 90-day period around national-level election votes along with a data set of tweets extracted using a keyword relevancy algorithm about the election during the same time period. We perform this analysis for the 2015 Nigeria and 2013 Kenya presidential elections.

We select the Kenyan and Nigerian elections because they are highly likely to include ethnic and religious grievances and contain varying levels of contentious activity, as described in detail in Section 3.1. Grievances, or collective emotional responses to perceived injustices, are theorized to decrease barriers to mobilization by focusing frustrations into aggression (Gurr (1968)). We expect the emotions generated by underlying grievances to be expressed in election-relevant communication on social media with more volatility over time than the inequalities that generate them. We capture the emotive content as contentious topics discussed before, during and after the vote to activate salient group divisions underlying political cleavages expecting to win or lose, fairly or unfairly, with the incumbent or opposition. Through topics and keywords, we test whether variation in this communication may provide a leading indicator for violent and nonviolent events. We find the mutually exciting HBTM accounts for topic spatial and temporal contagion of risk between conflict events of different types and Tweets related to the elections on social media.

The output of the model is an estimated intensity of events along with identified event cascades across topic space and time. Our approach is related to stochastic declustering (Zhuang, Ogata and Vere-Jones (2002)) of earthquake catalogs; however, our objective is to detect clusters and measure mutual excitation, whereas in seismology the goal is to remove clusters to obtain an estimate of the intensity of an earthquake's main shock. Our general methodology proceeds by viewing the point process as a branching process and then uses expectation-maximization (EM) to estimate the causal dependency of events and detect cascades of related events.

Previous studies on Hawkes models of social media have utilized temporal point processes (Zhao et al. (2015), Simma and Jordan (2012)), Latent Dirichlet Allocation (LDA) (Lai et al. (2016)) and Correlated Topic Model (CTM)-based point processes (He et al. (2015)). Dirichlet-Hawkes process is introduced in Du et al. (2015), Xu and Zha (2017) for modeling

social media and large-scale real-time burst modeling on Twitter is considered in Xie et al. (2016). In Farajtabar et al. (2017), a joint model for information diffusion and an evolving network is developed. In Dutta et al. (2020), the authors use Hawkes topic modeling for detecting fake retweeters and in Zheng et al. (2019) the authors use Hawkes topic modeling to analyze the dark web. Latent influencers are modeled in Tan, Rao and Neville (2018) using an Indian buffet Hawkes process. For a review of point process modeling of social diffusion, see Kim, Paine and Jurdak (2020).

Our work here considers a novel application, namely, detecting mutual excitation between online Twitter activity and real-world conflict and political instability events. The model we introduce is unique both in terms of the mutually-exciting component of the model capturing effects across heterogeneous data sets and the binomial topic model that facilitates a fast, easy-to-implement EM algorithm for inference while still outperforming the LDA Hawkes process used in Lai et al. (2016). Overall, the contributions of this work are as follows:

- We analyze a novel data set of election related tweets along with event data on protests, violence and territory change and estimate the casual influence across social media and conflict.
- We introduce a novel mutually-exciting Hawkes Binomial Topic Model. The mutual excitation component is novel for merging heterogeneous event-text data sets and estimating causality across them. The HBTM is novel as a model for cooccurrence of words between parent-child events in the Hawkes process. We show improved coherence of HBTM over LDA.

The outline of the paper is as follows. In Section 2 we introduce the Hawkes Topic Binomial Model and its extension to mutual excitation. Section 3 describes the Nigerian and Kenyan elections, the conflict event data and the methods used to retrieve and identify relevant Tweets. In Section 4 we present inference results on synthetic data and then study the detection of event cascades across heterogeneous event types and the overall causal structure of the estimated model applied to the social media and ACLED datasets. We assess the model's ability to forecast the level of conflict one day ahead against point process models with only univariate data as input and models without topic marks. In Section 5 we discuss the significance of our work along with future research directions.

**2. Hawkes binomial topic model.** Consider a marked Hawkes process with intensity  $\lambda(t, \vec{m})$  determined by

$$(1) \quad \lambda(t, \vec{m}) = \mu J_0(\vec{m} | p_0) + \sum_{t > t_i} \theta \omega e^{-\omega(t-t_i)} J_1(\vec{m}, \vec{m}_i | p_{\text{off}}, p_{\text{on}}).$$

Here, events at time  $t_i$  are associated with a mark  $\vec{m}_i$ , a vector of size  $W$  and the number of words in the overall dictionary across events. The binary variables indicate whether each word is present or absent in the event at time  $t_i$ . Spontaneous events occur according to a Poisson process with rate  $\mu$ . The mark vector of spontaneous events is determined by

$$(2) \quad J_0(\vec{m} | p_0) = p_0^{\sum_{j=1}^W m_j} (1 - p_0)^{W - \sum_{j=1}^W m_j}$$

which is the product of  $W$  independent Bernoulli random variables with parameters  $p_0$ .

The parameter  $\theta$  determines the expected number of events triggered by each event, and the expected waiting time between a parent-daughter event pair is given by  $\omega^{-1}$ . The mark of a daughter event is determined by two independent Bernoulli processes. Each word absent, or "turned off," in the parent bag of words is added to the bag of words of the child event with probability  $p_{\text{on}}$ . Each word present in the parent bag of words is deleted with probability  $p_{\text{off}}$ . Thus,  $J_1$  is given by

$$(3) \quad J_1(\vec{m}, \vec{m}_i | p_{\text{off}}, p_{\text{on}}) = p_{\text{on}}^{W_1^{\vec{m}, \vec{m}_i}} (1 - p_{\text{on}})^{W_2^{\vec{m}, \vec{m}_i}} p_{\text{off}}^{W_3^{\vec{m}, \vec{m}_i}} (1 - p_{\text{off}})^{W_4^{\vec{m}, \vec{m}_i}},$$

where  $W_1^{\vec{m}, \vec{m}_i}$  is the number of words present in the child vector and absent in the parent vector,  $W_2^{\vec{m}, \vec{m}_i}$  is the number of words absent in both vectors,  $W_3^{\vec{m}, \vec{m}_i}$  is the number of words in the parent vector absent in the child vector and  $W_4^{\vec{m}, \vec{m}_i}$  is the number of words present in both vectors.

While this model specification may not be appropriate for all dynamic topic analyses, we believe it works well when documents are short sentences or bags of words, as is the case with Tweets and event descriptions. After removing stop words we restrict the dictionary to the  $W$  most frequent words on the order of several hundred most frequent words across events.

2.1. *Estimation via expectation–maximization.* The model given by equation (1) can be viewed as a branching process (Mohler et al. (2011), Veen and Schoenberg (2008)) where events occur according to a stationary Poisson process  $\mu J_0(\vec{m} | p_0)$  and then each event  $i$  generates a Poisson process with intensity  $\theta \omega e^{-\omega(t-t_i)} J_1(\vec{m}, \vec{m}_i | p_{\text{off}}, p_{\text{on}})$ . Let  $u_{ij} = 1$  when event  $i$  is the direct offspring of event  $j$  and 0 otherwise and  $u_i^b = 1$  when event  $i$  is a “background” or “spontaneous” event generated by the background Poisson process (and 0 otherwise). Given knowledge of  $u_{ij}$ , the estimation problem decouples into several independent Poisson estimation problems. However, because  $u_{ij}$  is unknown, we introduce a matrix  $p_{ij}$  representing the probability that event  $j$  triggered event  $i$  and a vector  $p_i^b$  representing the probability that event  $i$  is a background event. Expectation–maximization inference then proceeds by iterating between the E-step:

$$(4) \quad p_{ij} = \frac{\theta \omega e^{-\omega(t_i-t_j)} J_1(\vec{m}_i, \vec{m}_j | p_{\text{off}}, p_{\text{on}})}{\lambda(t_i, \vec{m}_i)},$$

$$(5) \quad p_i^b = \frac{\mu J_0(\vec{m}_i | p_0)}{\lambda(t_i, \vec{m}_i)},$$

and the M-step:

$$(6) \quad \omega = \frac{\sum_{t_i > t_j} p_{ij}}{\sum_{t_i > t_j} p_{ij} (t_i - t_j)},$$

$$(7) \quad \mu = \frac{\sum_{i=1}^N p_i^b}{T},$$

$$(8) \quad \theta = \frac{\sum_{t_i > t_j} p_{ij}}{N},$$

$$(9) \quad p_0 = \frac{\sum_{i=1}^N p_i^b \|\vec{m}_i\|_1 / W}{\sum_{i=1}^N p_i^b},$$

$$(10) \quad p_{\text{on}} = \frac{\sum_{t_i > t_j} p_{ij} W_1^{\vec{m}_i, \vec{m}_j} / (W_1^{\vec{m}_i, \vec{m}_j} + W_2^{\vec{m}_i, \vec{m}_j})}{\sum_{t_i > t_j} p_{ij}},$$

$$(11) \quad p_{\text{off}} = \frac{\sum_{t_i > t_j} p_{ij} W_3^{\vec{m}_i, \vec{m}_j} / (W_3^{\vec{m}_i, \vec{m}_j} + W_4^{\vec{m}_i, \vec{m}_j})}{\sum_{t_i > t_j} p_{ij}}.$$

Here,  $N$  is the number of events in the data set, and  $T$  is the length of the time window. We note that the choice of the triggering kernel allows the parameters to be determined in the M-step via weighted sample mean estimators (Mohler (2014)). In equations (4)–(11) we have made the approximation that the integrals of the triggering kernel defined over the observation window can be replaced by integrals over the entire space (Mohler (2014)).

2.2. *Mutually exciting HBTM.* Next, we introduce an extension of HBTM allowing for mutual excitation across heterogenous data sets. For example, in Section 4 we model the coupled system of ACLED events with the subcategories protests, violence against civilians and territory change along with tweets from the 90-day period around an election yielding four total event categories. Twitter and conflict activities operate on different timescales, so we wish to allow for different Hawkes parameter values for both self- and mutual excitation.

For this purpose we add a second mark variable,  $s = 1, \dots, S$ , indicating the category of each event out of  $S$  categories. The conditional intensity is then determined by

$$(12) \quad \lambda_s(t, \vec{m}) = \mu_s J_0(\vec{m} | p_0^s) + \sum_{t > t_i} \theta_{ss_i} \omega_{ss_i} e^{-\omega_{ss_i}(t-t_i)} J_1(\vec{m}, \vec{m}_i | p_{\text{off}}^{ss_i}, p_{\text{on}}^{ss_i}).$$

Here, events of type  $s$  occur either according to a temporally stationary Poisson process with intensity  $\mu_s J_0(\vec{m} | p_0^s)$  or may be triggered instead by a previous event in the history of the process. The model parameters  $\mu_s$ ,  $p_0^s$ ,  $\theta_{sv}$ ,  $\omega_{sv}$ ,  $p_{\text{off}}^{sv}$  and  $p_{\text{on}}^{sv}$  now depend on  $s = 1, \dots, S$  and  $v = 1, \dots, S$  the categories of the child and parent events respectively. The EM algorithm for equation (12) is analogous to equations (4)–(11). Here, the only difference is that the summations over  $t_i > t_j$  for parameter with index  $s, v$  are restricted to  $i, j$  such that  $s_i = s, s_j = v$ .

**3. Cases, data, and methods for retrieval.** In this section we describe our case selection of the 2013 Kenyan and 2015 Nigerian elections and the conflict event data from ACLED. We then summarize our data collection process for retrieving Tweets that contain politically relevant communication. Expressions of grievance may be captured in this political content, thereby coupling social media messages to violent or contentious events.

3.1. *Case selection.* To investigate causation between social media and violent political events, we must select cases that exhibit two properties: high levels of social media use, and a variety of events with varying degrees of violence. Sub-Saharan Africa presents several cases that satisfy these constraints and has wide variation in infrastructure, social media use and political violence, so we focus our research on this geographical area. To capture political discourse, we further concentrate our work on timeframes surrounding election events in sub-Saharan Africa, as these events generally promote political discussion and increase the overall volume of discourse online. For this specific research, we study the 2015 Nigerian presidential election and the 2013 Kenyan presidential election, as both countries have substantial Internet-using populations and numerous relevant events. Regarding Internet use, 45.5% of Nigerians and 33.5% of Kenyans report using the Internet at least a few times a month, a few times a week or daily (Afrobarometer, Round 6, 2014–2015 ()). Based on qualitative examination, we expect the existence of underlying religious grievances in Nigeria and ethnic grievances in Kenya to impact variation in the number of violent and nonviolent events over time and space. Nigeria's 2015 election contains 565 events, and Kenya's 2013 election contains 281.

Below, we provide a brief summary of these elections to provide context for this study.

3.1.1. *2015 Nigerian presidential election.* On March 28, 2015, 42% of Nigeria turned out to vote in Nigeria's presidential election, pitting incumbent Goodluck Jonathan against Muhammadu Buhari, a retired general from the North with a strong anticorruption platform. This election was a near-identical reenactment of the 2011 electoral showdown between these two candidates, but this time, Buhari succeeded in unseating Jonathan. This opposition win made history in Nigeria as the first time an incumbent was defeated and peacefully stepped down (Owen and Usman (2015)). Despite the peaceful transition, the election was emotionally charged on both sides as the country battled security issues and persistent accusations

of electoral manipulation. State forces were too occupied with Boko Haram to ensure safe voting, resulting in low voter turnout, likely owing to anticipated violence. The incumbent administration also postponed the election, citing the security situation, serving only to inflame accusations of chicanery. The opposition accused the incumbent of manipulating the timing to advantage his party over the challenger. The mixture of inaction and dysfunction on the security question likely led many to believe that Buhari was a better choice than incumbent Jonathan. In Nigeria, while voter preferences tend to fall along ethnic and religious lines, parties cannot win without forming alliances and gaining support from outside their dominant religious or ethnic groups. As such, enough southern Christians voted for Buhari to secure a solid victory.

3.1.2. *2013 Kenyan presidential election.* On March 4, 2013, 86% of Kenyans turned out to vote in the presidential election, the first since catastrophic levels of violence occurred in 2007. In the previous election in 2007, at least 1133 people perished, and 700,000 people were displaced (Cheeseman, Lynch and Willis (2014)). In 2013, however, Kenya's ethnic grievances, which were responsible for violence in 2007, were dampened through considerable efforts at preventing violence, including a joint ticket between ethnic rivals Uhuru Kenyatta and William Ruto. Kenyatta, the incumbent from 2007's election, and former-rival-turned-running-mate William Ruto united to form the new umbrella Jubilee Coalition. While the International Criminal Court had individually indicted both Kenyatta and Ruto for their roles in the 2007 electoral violence, their alliance mitigated potential violence by aligning the rivaling Kalenjin and Kikuyu communities at the heart of the 2007 violence. Running against Kenyatta and Ruto was Raila Odinga and running mate Kalonzo Musyoka under the umbrella Coalition for Reforms and Democracy. All candidates regularly reiterated their commitment to peace, media outlets were trained in "conflict-sensitive reporting" in order to avoid inflaming tensions and security forces were heavily deployed in sensitive areas (Cheeseman, Lynch and Willis (2014)). This new election was considered free, fair and peaceful.

3.2. *Overview of conflict event data.* ACLED contains the broadest range of diverse event types of data sets that track conflict events (Eck (2012)), focusing on civil and communal conflicts, violence against civilians, remote violence and riots and protests (Raleigh et al. (2010)). The ACLED project codes the dates and locations of all reported protests and political violence within periods of instability, civil war and regime breakdown, capturing episodic activity, such as potential precursor events. Actors include political agents, including governments, rebels, militias, communal groups, political parties, rioters, protesters and civilians. From these event types we utilize protests, violence against civilians and transfer of territory to cover the broad spectrum of activity. Using the ACLED's date, location, event type and event description text, we create a unique data set by merging the observations with Tweets for the 90-day period around the election. These Tweets are curated using the following techniques to retrieve information, for query expansion and for identification of relevance within the corpus and metadata.

3.3. *Social media data collection.* To study the connection between social media and conflict, we must collect large volumes of relevant social media. The Twitter platform is ideal for this task, as its content is open and publicly available by default (in contrast to Facebook or Snapchat which restrict access to preselected individuals by default). Nigeria and Kenya also are in the top 25% of countries with presence on Twitter in terms of Tweets with geospatial location tags, called "geotags," and communicate primarily in English on Twitter. This section details our process for retrieving relevant content from Twitter's platform, mitigating biases in this content and adapting to "big data"-scale data sets. At a high level, this retrieval process consists of:



1. Human-expert query generation,
2. Computationally assisted, multidimensional query expansion,
3. Data reseller-backed retrieval, and
4. At-scale data sanitization.

3.3.1. *Expert query generation.* To prime our data collection pipeline, we ask subject-matter experts who have studied the Nigerian and Kenyan elections to generate semistructured descriptions of each election. These descriptions include collections of keywords, phrases and individuals for which we can search in a social media data set. Search terms could be lists of individuals involved, such as leaders, candidates, journalists and activists; associated organizations, supporting constituencies and ideologies; issues or drivers leading to contentious actions; and actions or events in the relevant time period, such as protests, rallies or attacks, court cases, voter registration problems, candidate or party announcements, scandals or political violence. Experts then distill these summaries into topics and subtopics and refine their descriptions through an iterative search process that uses both social and traditional media sources.

3.3.2. *Multidimensional query expansion.* Though experts can generate high-quality event descriptions, the Internet's varied populations and communities may express thoughts or refer to the same concepts in many different ways. Therefore, we assume event descriptions are incomplete and may omit a subset of tokens some populations may use to refer to a particular entity or concept. To address this issue, we leverage a large, undirected sample of Twitter data to expand these expert-generated queries and capture relevant tokens in a data-driven manner. This sample is collected from Twitter's public sample stream which publishes 1% of all Tweets posted to the Twitter platform. Furthermore, "subtweeting," or referencing an event or entity without directly mentioning it, complicates retrieval, so we also expand queries using social and spatial dimensions (the spatial dimension is expanded via the data reseller).

Textual expansion is performed by querying our undirected Twitter sample for content that matches expert-generated queries. We then identify keywords whose frequencies in the set of matching messages are much higher than in the general sample by ranking tokens by their Kullback–Leibler (KL) divergence (Kullback and Leibler (1951)). Keywords with a KL divergence exhibit a strong cooccurrence connection with the original query but are rare in the underlying sample. These keywords are useful for identifying potentially missing keywords for the event particular to Twitter and its communities' differing vocabularies (Gouws et al. (2011)). Specifically, we find all messages in the random sample that match the original query, tokenize these messages into bags of words, rank by KL divergence score and ask our experts to determine which keywords to add to the original query.

To expand queries along the social dimension, we identify socially-relevant users by converting the retweet and mention activity in a sample of relevant content into a directed graph of interactions. This relevant sample is generated from Tweets matching the original expert queries described above. Research shows highly followed or retweeted users are often not the most influential users (Cha et al. (2010)), so we follow Kwak et al. and use a version of Google's PageRank algorithm to calculate an "authority" score for users in this interaction graph (Kwak et al. (2010)). Experts are then presented with a list of accounts ranked by their authority score and asked which accounts should be added to the original queries.

3.3.3. *Information retrieval from data resellers.* After expanding experts' event descriptions and queries, we search Twitter's full historical archive to acquire a large data set of Tweets for each election. However, Twitter's historical archive is not freely available, so we

contract with Gnip, Twitter’s primary data reseller, to access this data. We use Gnip’s native support for textual, social, spatial and temporal queries to search for relevant Tweets that match keywords/phrases and/or mentions are from relevant accounts or are posted from within the target country, all of which are restricted to 60 days prior and 30 days following the election date in each country.

Using this process, we retrieved 7.1 and 4.7 million Tweets for Nigeria and Kenya, respectively.

3.3.4. *At-scale data sanitization.* The previous steps in this data retrieval process ensure significant *recall* in our data; we therefore avoid omitting relevant content (high recall) at the cost of including irrelevant Tweets (low precision). However, to ensure high-quality modeling with HBTM, we must clean these data sets and remove content that is not relevant to the political contexts of these countries. Standard approaches have experts sanitize the data through review and ensuring relevance, but when dealing with several million social media messages, complete human review is intractable. To support sanitization at scale, we use a mixed-methods, iterative approach to train an automated system that can identify relevant Tweets.

We bootstrap this sanitization process by randomly sampling 300 Tweets that match textual, social and spatial queries, resulting in a set of approximately 900 Tweets. We then perform several rounds of relevance feedback in which pairs of expert analysts review these 300 tweets, labeling each as relevant, irrelevant, not English or undecidable. Labels are analyzed for inter-rater reliability, experts meet to discuss their labeling strategies and labeling is repeated until agreement between experts reaches a sufficient level of agreement, which we set as a Cohen’s  $\kappa > 0.61$ . Tweets agreed to be relevant and irrelevant are saved for classifier training, while undecided or non-English Tweets are discarded.

For Nigeria, our annotators achieved an agreement of  $\kappa = 0.75$  and agreed on 836 labeled Tweets, 495 of which were labeled as relevant to the Nigerian election. In Kenya, agreement reached  $\kappa = 0.69$  and 688 usable tweets, 321 of which were labeled as relevant.

To train a classifier to identify relevance from these data sets, we featurize each Tweet into a bag of words and train a set of Gradient Boosted Trees (GBTs)<sup>1</sup> on these feature vectors and their relevance labels. We train separate GBTs for each election event as well. We then apply these trained classifiers to the full set of Tweets for each event and store all instances classified as relevant. We then randomly sample 1000 Tweets from these potentially relevant Tweets and pass them through a second round of relevance feedback to ensure our classifiers are performing well. Labels from this second round of relevance feedback are incorporated into our relevance models as well.

In this second round of assessment, our experts agreed 93% of the Nigerian Tweets the classifier identified as relevant were correct. For Kenya, however, our experts only found 49% of the 1000 sampled Tweets to be relevant. These results suggest Nigeria has a large quantity of relevant data; however, Kenya’s data is either noisy, or our models cannot capture political discourse using only textual data.

To determine whether our focus on textual content was adequate for relevance classification, we ran an experiment that trained classifiers on the other dimensions present in Tweets as well: the standard bag-of-words model, a bag-of-mentions model (where tokens are replaced by Twitter users mentioned in a Tweet), bag-of-locations (using self-reported user locations) and temporal classification as well (i.e., how close is the Tweet to the event). For Nigeria and Kenya we generated precision-recall curves for each set of features using 30 rounds of

<sup>1</sup>While other classifiers, such as support vector machines and random forests, perform equivalently (we tested these, and they achieved similar performance scores), GBTs are easily distributed across cluster systems.



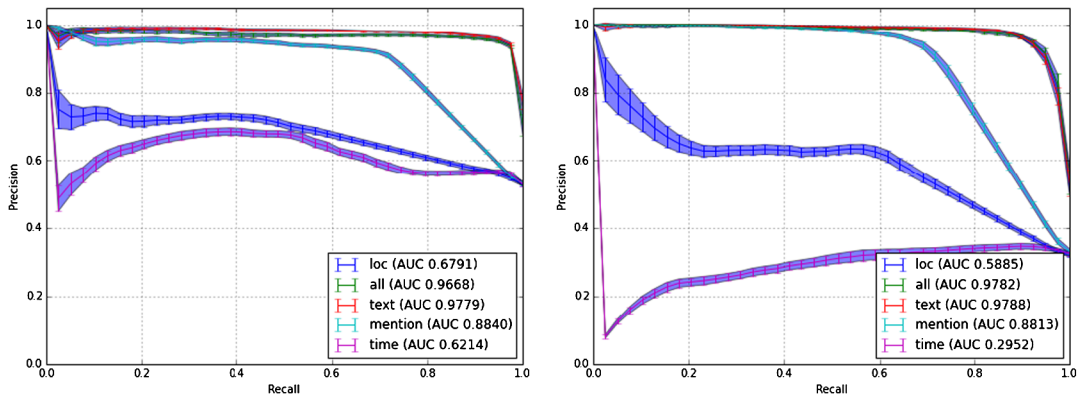


FIG. 2. Precision-recall curves for various dimensions in Nigeria (left) and Kenya (right).

random sampling in which 25% of our labels were held out for testing, and a classifier was trained on the remaining 75%. Precision and recall were calculated across these 30 rounds, and we calculated the mean area under the curve (AUC) and 95% confidence intervals for each feature set and a union of all feature sets. Results, shown in Figure 2, show that textual features perform as well or better than all other features separately and together.

This sanitization process' final step is to restrict data to only those Tweets classified as relevant *and* are from social media users who participate in communities discussing these election events. We identify communities in Tweets by applying label propagation to the interaction graph described above and keep the three largest communities who reference these events. This step is primarily useful to remove references to elections outside the target countries, as elections in the United States and India use similar language and hashtags that could contaminate our data.

This process yielded 2.8 million Tweets relevant to Nigeria and 1.1 million Tweets relevant to Kenya.

## 4. Results.

4.1. *Simulated data.* We first test the HBTM EM estimation procedure by simulating several realizations of the HBTM process and assessing parameter recovery. In particular, we simulate a mutually exciting HBTM with two categories and parameters given by

$$\mu = [1, 2], \quad \theta = \begin{bmatrix} 0.5 & 0.1 \\ 0.2 & 0.7 \end{bmatrix}, \quad \omega = \begin{bmatrix} 0.1 & 1.0 \\ 0.05 & 0.9 \end{bmatrix},$$

$$p_{\text{on}} = \begin{bmatrix} 0.05 & 0.1 \\ 0.05 & 0.07 \end{bmatrix}, \quad p_{\text{off}} = \begin{bmatrix} 0.2 & 0.1 \\ 0.3 & 0.1 \end{bmatrix}, \quad p_0 = [0.1, 0.08] \quad \text{and} \quad W = 200.$$

In Figure 3 we plot the true value of the parameters along with a histogram of estimated parameters from 100 realizations of the point process where 10 iterations of EM were used in each simulation. Here, we find that all parameters are recovered with error on the order of or less than 10%. We have posted the Matlab code to replicate the simulations on Github (see the Supplementary Material (Mohler et al. (2020))).

4.2. *Coupled ACLED-Twitter data.* We next analyze the merged ACLED and Twitter data set of events occurring in Nigeria from February 2 through April 30, 2015, and Kenya from January 8, 2013, to April 2, 2013. In particular, we consider four categories of events in the given time periods: protests, violence against civilians, battles with territory change and

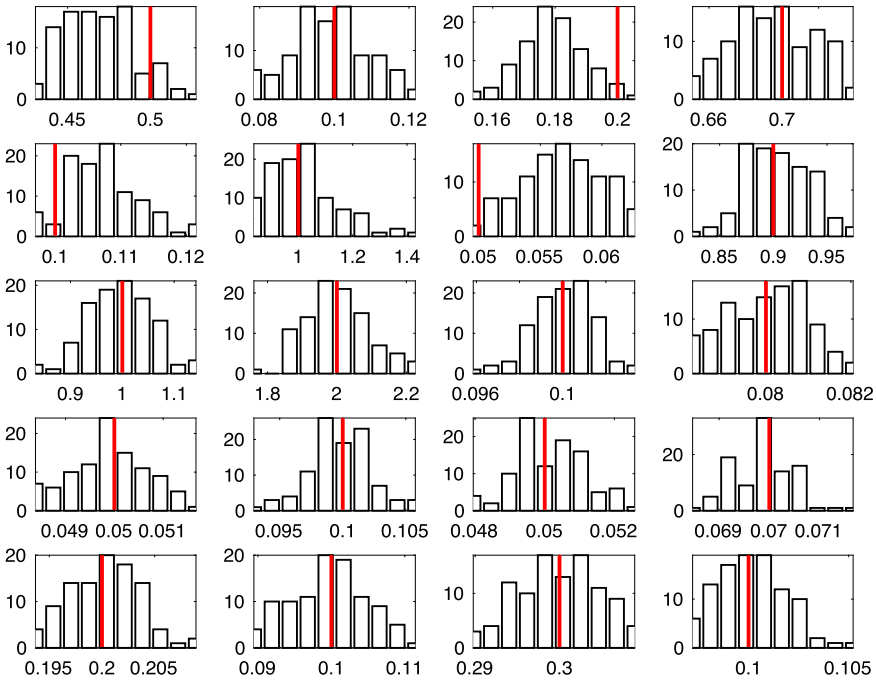


FIG. 3. EM estimates for 100 realizations of the mutually exciting HBTM. Red line indicates ground truth. Top row corresponds to  $\theta$ , second row corresponds to  $\omega$ , third row corresponds to  $\mu$  and  $p_0$ , fourth row corresponds to  $p_{on}$  and fifth row corresponds to  $p_{off}$ .

tweets relevant to the elections during the time periods. Given the quadratic cost of the EM algorithm, we subsample the Twitter data after the relevancy algorithm is applied to bring the size of the Twitter data down to  $O(10^4)$  events. In particular, the Kenya dataset has 6108 events, and the Nigeria dataset has 8412 events.

Using the expectation–maximization algorithm from Section 2, we estimate the mutually exciting HBTM and find that the EM algorithm converges within 10–20 iterations. We first investigate the extent of self- and mutual excitation across events of different categories by constructing a weighted, directed graph where the edge weights are determined by  $\theta_{ml}N_l$ , representing the number of direct offspring events of category  $m$  generated by events of type category  $l$ . We include the background rates as nodes on the network as well, letting the edge weights be determined by  $\mu_m T$ , representing the number of background events of type  $m$ .

Several patterns emerge upon inspection of the graph in Figure 4. First, in both Nigeria and Kenya Tweets are predominately generated through selfexcitation. The lack of influence of ACLED events on Twitter may be, in part, explained by the lack of hourly resolution of ACLED events and the fast time scale on which Tweet patterns evolve. Nigeria and Kenya graphs both have similar edge characteristics, with the exception of protest-territory mutual excitation. In Tables 1–3 we display the parameter estimates for the ME-HBTM corresponding to the graph in Figure 4. We obtain compute standard errors by repeated simulation of the process with the MLE parameters and computing bootstrapped estimates. In Table 3 we observe the productivity parameter values  $\theta$ . For example, we estimate that in Kenya and Nigeria one protest is generated by every 2000 election related tweets ( $\theta_{ij} = 0.0005$ ). Based on the standard errors, this value is statistically significant.

We next investigate the extent to which the model predicts daily risk of ACLED event activity in the case of the mutually exciting HBTM (ME-HBTM), HBTM without mutual excitation (four independent Hawkes processes with  $\theta$  diagonal) and a mutually exciting Hawkes process with no Topic mark component. For each day in the 90 day periods, we

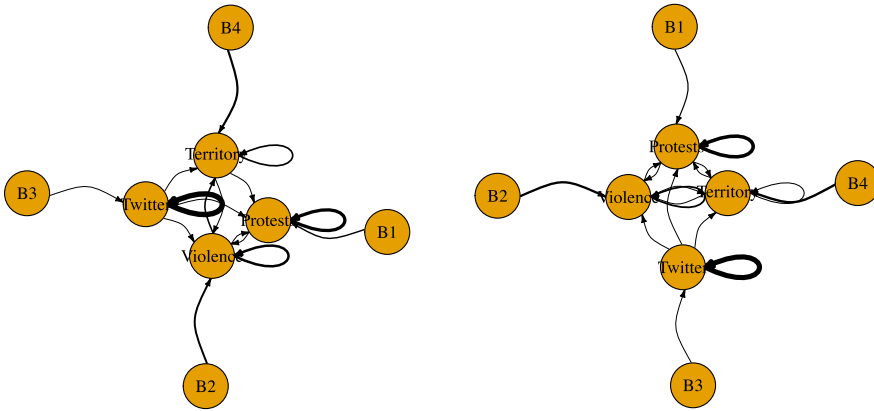


FIG. 4. Graphical representation of estimated point process model fit to the ACLED-Twitter event data set for Nigeria (left) and Kenya (right). Weights of the edges of the directed graph correspond to the fraction of events triggered across the edge. B1–B4 represent the background rate of the process.

average the intensity over the empirical mark distribution for that day and use the mean intensity to predict the number of events the following day. We then use the Normalized Gini Index (NGI) to assess each model’s performance (Frees, Meyers and Cummings (2011)). The NGI is a continuous analog to AUC measuring the ability of each model to correctly sort days by the number of events, where errors for high event days are penalized greater than low event days (NGI is normalized such that 1 indicates a perfect prediction).

In Tables 4 and 5 we display the NGI scores for the three models. In both Kenya and Nigeria, protests are most accurately predicted using the ME-HBTM, likely due to protests being the most relevant ACLED event to the elections and social media discussions. In Kenya, the HBTM selfexciting model performs best for violence and territory event types. In Nigeria, territory is most accurately predicted with a mutually exciting Hawkes process; however, the topic models perform better for protest and violence.

The EM branching process formulation of the problem requires that the branching structure of the process be estimated simultaneously with the model parameters. Event cascades can then be detected using the triggering probability matrix  $p_{ij}$  from Section 2. In particular, we sample from the probabilities and then merge all events into clusters that are ancestors.

In Figure 5, we compare the UCI coherence (Newman et al. (2010)) of the ME-HBTM model to the coherence of LDA (where the number of topics  $k$  for LDA is varied). Here, we see that the ME-HBTM model has improved coherence, a measure of the tendency of words to cooccur at a greater frequency if they are in the same topic. This can be explained by the ME-HBTM construction, where parent-child events tend to share the same words in the model. Another advantage of the ME-HBTM is that the number of topics is automatically selected.

In Figure 6 we plot several of the largest clusters for each of the elections. In Kenya we detect several large clusters of events around the election debates, the election itself,

TABLE 1  
Mean and standard error for  $\mu$  in ME-HBTM

	Protest	Violence	Twitter	Territory
Nigeria	0.053 +/- 0.004	0.047 +/- 0.003	0.435 +/- 0.026	0.032 +/- 0.002
Kenya	0.056 +/- 0.008	0.080 +/- 0.011	4.108 +/- 0.108	0.001 +/- 0.003

TABLE 2  
Mean and standard error for  $\omega$  in ME-HBTM

Nigeria	Protest	Violence	Twitter	Territory
Protest	0.306 +/- 0.278	0.215 +/- 1.375	0.271 +/- 0.744	0.3989 +/- 4.620
Violence	0.776 +/- 2.185	0.255 +/- 0.054	1.360 +/- 0.318	0.250 +/- 0.152
Twitter	0.229 +/- 0.622	18.483 +/- 1.113	0.818 +/- 0.005	1.846 +/- 0.583
Territory	0.401 +/- 0.511	0.252 +/- 1.544	1.481 +/- 0.523	1.498 +/- 0.284

Kenya	Protest	Violence	Twitter	Territory
Protest	0.359 +/- 19.014	1.097 +/- 0.154	11.378 +/- 0.846	0.414 +/- 1.157
Violence	0.865 +/- 0.402	0.261 +/- 0.215	0.556 +/- 0.052	0.265 +/- 0.671
Twitter	21.706 +/- 2.025	0.699 +/- 0.220	2.497 +/- 0.007	3.612 +/- 2.933
Territory	1.360 +/- 3.716	0.197 +/- 6.042	0.541 +/- 0.659	0.217 +/- 0.125

protests and a cluster corresponding to a supreme court petition regarding the validity of the election results. Three major topics arising in the Nigeria clusters center around the election, protests and Boko Haram attacks. The selfexcitation productivity parameter  $\theta_{mm}$  of Twitter events for the Nigeria data set is close to 1 ( $\approx 0.99$ ) yielding large, long-lasting clusters. For comparison, we display the results of LDA topic modeling in Figure 7. Unlike LDA, the HBTM is capable of detecting new clusters and automatically determines the number of clusters. In particular, there are over 100 topic-time clusters found in each data set by HTBM which may be viewed as microtopics in time compared to topics returned by LDA.

4.3. *Goodness of fit.* In equation (1) we make several assumptions on the triggering kernel, namely, that the kernel is separable in topic space and time and that the distribution in time is exponential. In this subsection we investigate the extent to which these assumptions may be valid. To do this, we sample from the estimated branching probabilities  $p_{ij}$  in equation (4). If the model is correctly specified, then sampled interevent data  $(t_i - t_j, m_i, m_j)$  will be realizations from the triggering kernel density. We can then assess separability and model fit based on this data.

First, we investigate separability in time and topic space. To reduce the dimension in mark space, we apply principal component analysis to the document-term matrix and focus on

TABLE 3  
Mean and standard error for  $\theta$  in ME-HBTM

Nigeria	Protest	Violence	Twitter	Territory
Protest	0.697 +/- 0.032	0.062 +/- 0.017	0.025 +/- 0.016	0.002 +/- 0.006
Violence	0.079 +/- 0.012	0.561 +/- 0.023	0.008 +/- 0.019	0.322 +/- 0.020
Twitter	0.0005 +/- 0.0003	0.0004 +/- 0.0001	0.994 +/- 0.004	0.0001 +/- 0.0001
Territory	0.129 +/- 0.013	0.174 +/- 0.019	0.009 +/- 0.039	0.434 +/- 0.029

Kenya	Protest	Violence	Twitter	Territory
Protest	0.751 +/- 0.027	0.073 +/- 0.012	0.027 +/- 0.067	0.085 +/- 0.009
Violence	0.065 +/- 0.017	0.439 +/- 0.020	0.002 +/- 0.093	0.079 +/- 0.009
Twitter	0.0005 +/- 0.00006	0.0005 +/- 0.00005	0.940 +/- 0.002	0.0006 +/- 0.00006
Territory	0.184 +/- 0.033	0.591 +/- 0.043	0.104 +/- 0.307	0.299 +/- 0.020

TABLE 4  
NGI scores for Kenya

	ME-HBTM	HBTM	ME-H
Protest	<b>0.3990</b>	0.3546	0.3558
Violence	0.2247	<b>0.2357</b>	0.1399
Territory	0.5798	<b>0.6134</b>	0.4958

TABLE 5  
NGI scores for Nigeria

	ME-HBTM	HBTM	ME-H
Protest	<b>0.3159</b>	0.2869	0.0853
Violence	<b>0.4191</b>	0.3700	0.2830
Territory	0.1549	0.0771	<b>0.2662</b>

the first principal component, which we denote by “PCA” in Figure 8. We then consider the interevent pairs  $(t_i - t_j, PCA_i - PCA_j)$  and conduct a Hoeffding test for dependence using the R package testforDEP. We find that, in the case of Kenya, we cannot reject the null hypothesis of independence, with a p-value of 0.08, though in the case of Nigeria we find some limited evidence of dependence among the two variables, with a p-value of 0.002. However, in Figure 8 we see that the correlation is low both in the case of Kenya (0.004) and Nigeria (0.022), and we do not find enough deviation from separability to warrant a more complex, nonseparable model specification.

Next, we investigate the goodness of fit of the exponential-in-time assumption for the triggering kernel. Again, we sample from the estimated branching probabilities  $p_{ij}$  to yield realizations  $t_i - t_j$  from the triggering kernel  $\omega \exp(-\omega t)$  (under the assumption that the model is correctly specified). We note that, in our multivariate Hawkes process model, there

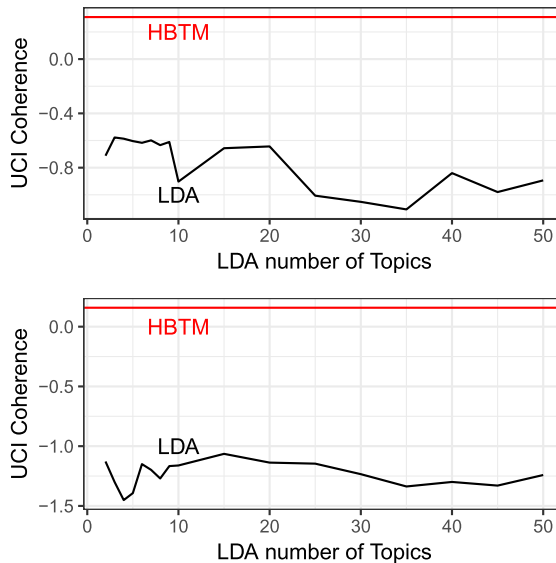


FIG. 5. UCI coherence of HBTM vs. LDA for Nigeria (top) and Kenya (bottom) datasets.

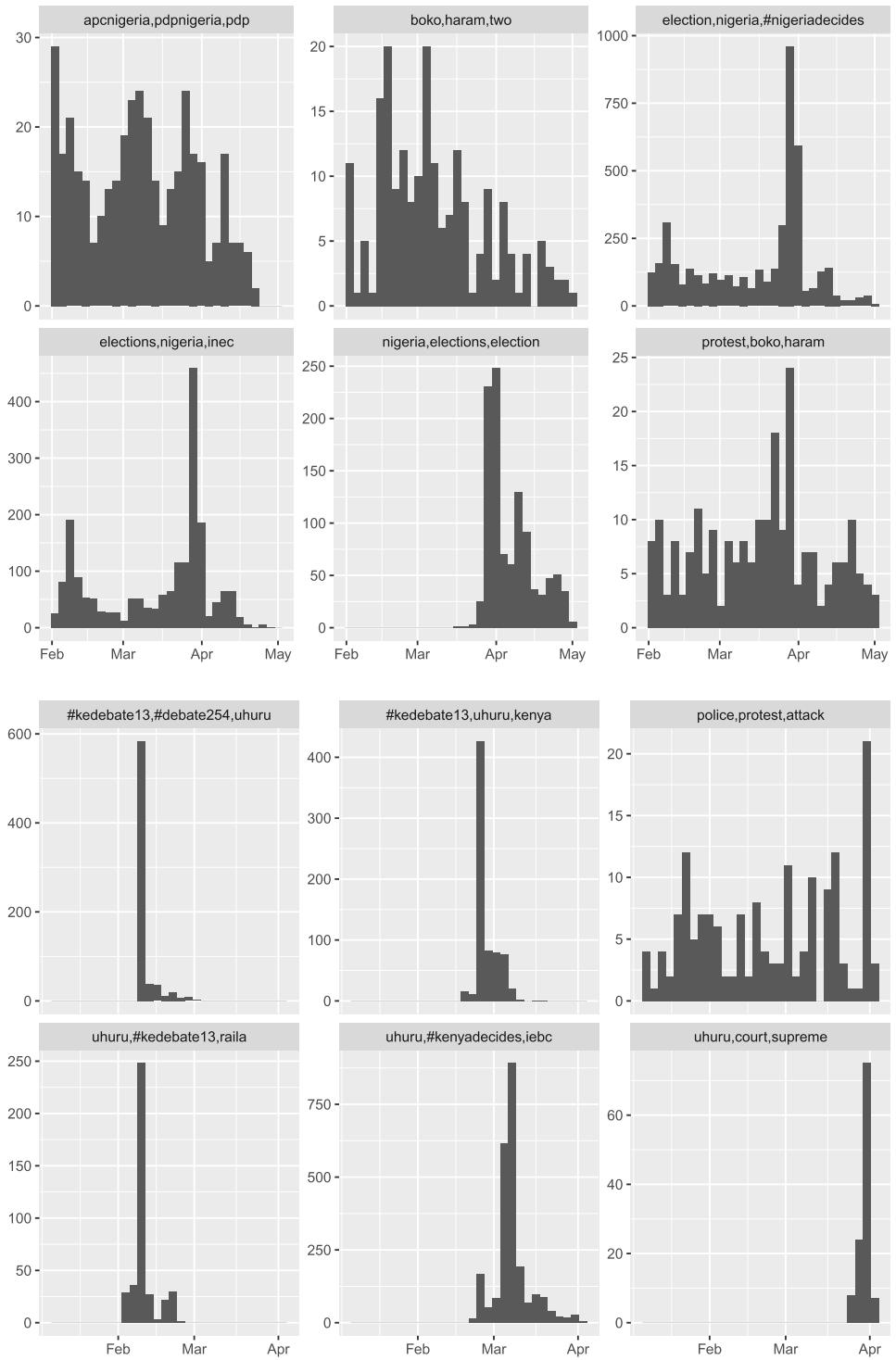


FIG. 6. Histograms of event times of HBTM clustered events along with the three most frequent keywords words in each cluster for Nigeria (top two rows) and Kenya (lower two rows).



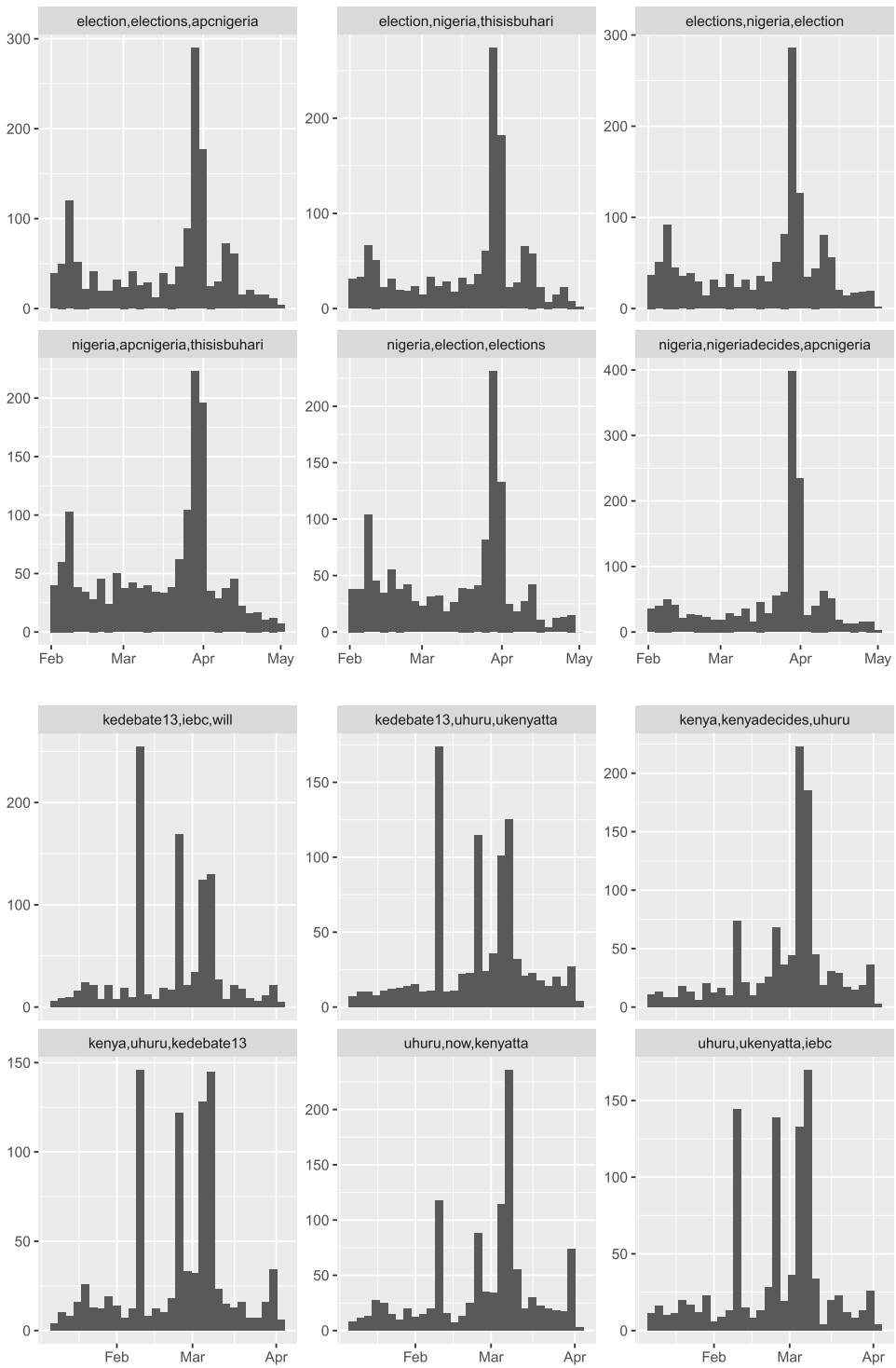


FIG. 7. Histograms of event times of LDA clustered events along with the three most frequent keywords words in each cluster for Nigeria (top two rows) and Kenya (lower two rows).

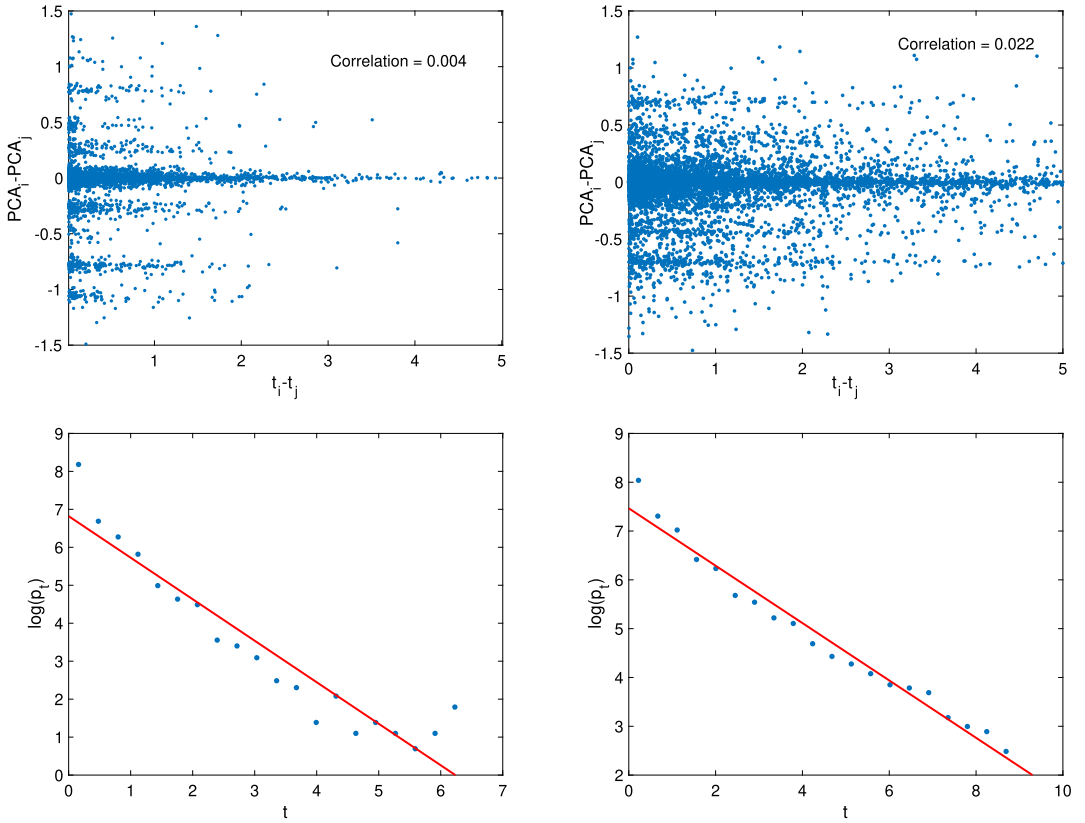


FIG. 8. Top: Scatter plot of the difference in mark first principal component versus difference in time of parent-offspring pairs for Kenya (left) and Nigeria (Right). Bottom: Best fit line to empirical histogram (log-scale) of sampled parent-offspring inter-event times  $t_i - t_j$  for Kenya (left) and Nigeria (right).

are 16 crossexcitation kernels and that the data set size is small,  $O(50)$ , for some of the category-category pairs. We therefore restrict our attention to parametric models rather than consider nonparametric kernels that may be more appropriate for larger data set sizes.

A common alternative choice in Hawkes processes to the exponential for the triggering kernel is a heavy tailed Pareto distribution, and we compute the AIC for these two competing models for the diagonal components (self-excitation) in Table 6. Here, we see that for protest, violence and territory the one parameter exponential is favored over the two parameter Pareto. However, for Twitter the Pareto distribution provides a better fit. Nonetheless, in Figure 8 we compare the empirical histogram of the interevent times for Twitter to a best fit exponential and note that the model provides a reasonable fit to the data.

TABLE 6  
*Model fit of exponential vs. pareto diagonal (selfexciting) triggering kernels*

	Exp AIC	Pareto AIC
Protest	348.1	350.1
Violence	209.5	210.8
Twitter	842.7	-510.3
Territory	60.7	62.7

**5. Significance and impact.** While previous point process studies of conflict and terrorism have modeled single event types, often in time only, we presented a framework that extends previous work to include crossexcitation within conflict and social media effects coupled with a topic model that allows for the incorporation of event descriptions. We believe that such a framework could find application not only in conflict studies but also in other scenarios where high-dimensional information accompanies each event in a point process. For example, in the case of crime records databases often event descriptions are available but seldom used in space-time models of crime. The methodology outlined in this paper could be used to incorporate that information and to solve crime linkage problems where the goal is to match crimes committed by the same offender (Porter (2016)).

Our methodology also provides a tool for merging and summarizing coupled conflict-social media event data sets. We create and implement a framework to identify relevant conversations around elections thought to heighten tensions that expose underlying grievances between groups that have been and will be effected by those voted into power, and the perception of whether or not the electoral process was free and fair. We create an unbiased corpus by expanding our expert query with data-driven textual, social and spatial dimensions that may have been omitted and then, using machine learners, trained on human annotation to increase precision at scale. We couple data from these Tweets with conflict events and perform event cascade detection where groups of events are identified that are both temporally, categorically (mark type) and topically related through the estimated branching process.

We illustrated this process using data from the time period around the 2015 Nigerian election and the 2013 Kenyan election, summarizing the regional instability and national conversations on the election during the time period. The ACLED event data captures all conflict or violent events, not just election-related violence. However, the model indicates that discourse on Twitter specifically related to the election likely drives protests more than violence and territory changes, reflected in the model's assessed capabilities to capture a leading indicator for events. This is a significant finding for the potential of national-level election conversations on social media to translate existing grievances into contentious action and or conflict.

In the future it may be worthwhile to consider extensions of the HBTM model framework. In its current form the model ignores the potential dependency among various words, for example, there is some probability that a word in a parent event is replaced by a synonym in the offspring event. Furthermore, there may be situations where a nonseparable and/or nonparametric triggering kernel is appropriate. These will be directions for future research.

**Acknowledgments.** This research was supported by NSF Grants SCC-1737585, ATD-1737996 and SES-1343123.

## SUPPLEMENTARY MATERIAL

**Source code for “Hawkes binomial topic model with applications to coupled conflict-Twitter data.”** (DOI: [10.1214/20-AOAS1352SUPP](https://doi.org/10.1214/20-AOAS1352SUPP); .zip). Matlab code for synthetic data experiment.

## REFERENCES

- AFROBAROMETER. Kenya and Nigeria, Round 6, 2014–2015.
- BELLIN, E. (2012). Reconsidering the robustness of authoritarianism in the Middle East: Lessons from the Arab Spring. *Comp. Polit.* **44** 127–149.
- BHAVNANI, R. and DONNAY, K. (2012). Here's looking at you: The Arab Spring and violence in Gaza, Israel and the West Bank. *Swiss Pol. Sci. Rev.* **18** 124–131.
- BUNTAIN, C., GOLBECK, J., LIU, B. and LAFREE, G. (2016). Evaluating public response to the Boston marathon bombing and other acts of terrorism through Twitter. In *Tenth International AAAI Conference on Web and Social Media*.

- CHA, M., HADDADI, H., BENEVENUTO, F. and GUMMADI, K. P. (2010). Measuring user influence in Twitter: The million follower fallacy. In *Fourth International AAAI Conference on Weblogs and Social Media*. Available at <http://aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1538/0>.
- CHEESEMAN, N., LYNCH, G. and WILLIS, J. (2014). Democracy and its discontents: Understanding Kenya's 2013 elections. *J. East. Afr. Stud.* **8** 2–24.
- DU, N., FARAJTABAR, M., AHMED, A., SMOLA, A. J. and SONG, L. (2015). Dirichlet–Hawkes processes with applications to clustering continuous-time document streams. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 219–228. ACM, New York.
- DUTTA, H. S., DUTTA, V. R., ADHIKARY, A. and CHAKRABORTY, T. (2020). Hawkeseye: Detecting fake retweeters using Hawkes process and topic modeling. *IEEE Trans. Inform. Forensics Secur.* **15** 2667–2678.
- ECK, K. (2012). In data we trust? A comparison of UCDP GED and ACLED conflict events datasets. *Coop. Confl.* **47** 124–141.
- FARAJTABAR, M., WANG, Y., GOMEZ-RODRIGUEZ, M., LI, S., ZHA, H. and SONG, L. (2017). COEVOLVE: A joint point process model for information diffusion and network evolution. *J. Mach. Learn. Res.* **18** Paper No. 41, 49. [MR3655306](https://doi.org/10.1145/31655306)
- FREES, E. W., MEYERS, G. and CUMMINGS, A. D. (2011). Summarizing insurance scores using a Gini index. *J. Amer. Statist. Assoc.* **106** 1085–1098. [MR2894766 https://doi.org/10.1198/jasa.2011.tm10506](https://doi.org/10.1198/jasa.2011.tm10506)
- GOUWS, S., METZLER, D., CAI, C., HOVY, E. and REY, M. (2011). Contextual bearing on linguistic variation in social media. In *Proceedings of the Workshop of Language in Social Media*.
- GURR, T. (1968). Psychological factors in civil violence. *World Polit.* **20** 245–278.
- HE, X., REKATSINAS, T., FOULDS, J., GETOOR, L. and HAWKESTOPIC, Y. L. (2015). A joint model for network inference and topic modeling from text-based cascades. In *Proc. 32nd Intl. Conf. on Machine Learning*.
- KIM, M., PAINI, D. and JURDAK, R. (2020). Real-world diffusion dynamics based on point process approaches: A review. *Artif. Intell. Rev.* **53** 321–350.
- KULLBACK, S. and LEIBLER, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.* **22** 79–86. [MR0039968 https://doi.org/10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694)
- KWAK, H., LEE, C., PARK, H. and MOON, S. (2010). What is Twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web* 591–600. ACM, New York.
- LAI, E. L., MOYER, D., YUAN, B., FOX, E., HUNTER, B., BERTOZZI, A. L. and BRANTINGHAM, P. J. (2016). Topic time series analysis of microblogs. *IMA J. Appl. Math.* **81** 409–431. [MR3564661 https://doi.org/10.1093/imamat/hxw025](https://doi.org/10.1093/imamat/hxw025)
- MOHLER, G. (2013). Modeling and estimation of multi-source clustering in crime and security data. *Ann. Appl. Stat.* **7** 1525–1539. [MR3127957 https://doi.org/10.1214/13-AOAS647](https://doi.org/10.1214/13-AOAS647)
- MOHLER, G. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *Int. J. Forecast.* **30** 491–497.
- MOHLER, G. O., SHORT, M. B., BRANTINGHAM, P. J., SCHOENBERG, F. P. and TITA, G. E. (2011). Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* **106** 100–108. [MR2816705 https://doi.org/10.1198/jasa.2011.ap09546](https://doi.org/10.1198/jasa.2011.ap09546)
- MOHLER, G., MCGRATH, E., BUNTAIN, C. and LAFREE, G. (2020). Supplement to “Hawkes binomial topic model with applications to coupled conflict-Twitter data.” <https://doi.org/10.1214/20-AOAS1352SUPP>
- NEWMAN, D., LAU, J. H., GRIESER, K. and BALDWIN, T. (2010). Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* 100–108. Association for Computational Linguistics, Stroudsburg.
- OWEN, O. and USMAN, Z. (2015). Why goodluck jonathan lost the nigerian presidential election of 2015. *Afr. Aff.* adv037.
- PORTER, M. D. (2016). A statistical approach to crime linkage. *Amer. Statist.* **70** 152–165. [MR3511045 https://doi.org/10.1080/00031305.2015.1123185](https://doi.org/10.1080/00031305.2015.1123185)
- PORTER, M. D. and WHITE, G. (2012). Self-exciting hurdle models for terrorist activity. *Ann. Appl. Stat.* **6** 106–124. [MR2951531 https://doi.org/10.1214/11-AOAS513](https://doi.org/10.1214/11-AOAS513)
- RALEIGH, C., LINKE, A., HEGRE, H. and KARLSEN, J. (2010). Introducing ACLED-armed conflict location and event data. *J. Peace Res.* **47** 651–660.
- RENNICK, S. A. (2013). Personal grievance sharing, frame alignment, and hybrid organisational structures: The role of social media in North Africa's 2011 uprisings. *J. Contemp. Afr. Stud.* **31** 156–174.
- SIMMA, A. and JORDAN, M. I. (2012). Modeling events with cascades of Poisson processes. Preprint. Available at [arXiv:1203.3516](https://arxiv.org/abs/1203.3516).
- TAN, X., RAO, V. and NEVILLE, J. (2018). The Indian Buffet Hawkes process to model evolving latent influences. In *UAI* 795–804.
- VEEN, A. and SCHOENBERG, F. P. (2008). Estimation of space-time branching process models in seismology using an EM-type algorithm. *J. Amer. Statist. Assoc.* **103** 614–624. [MR2523998 https://doi.org/10.1198/016214508000000148](https://doi.org/10.1198/016214508000000148)

- XIE, W., ZHU, F., JIANG, J., LIM, E.-P. and TOPICKSKETCH, K. W. (2016). Real-time bursty topic detection from Twitter. *IEEE Trans. Knowl. Data Eng.* **28** 2216–2229.
- XU, H. and ZHA, H. (2017). A Dirichlet mixture model of Hawkes processes for event sequence clustering. In *Advances in Neural Information Processing Systems* 1354–1363.
- ZAMMIT-MANGION, A., DEWAR, M., KADIRKAMANATHAN, V. and SANGUINETTI, G. (2012). Point process modelling of the Afghan war diary. *Proc. Natl. Acad. Sci. USA* **109** 12414–12419.
- ZAMMIT-MANGION, A., DEWAR, M., KADIRKAMANATHAN, V., FLESKEN, A. and SANGUINETTI, G. (2013). Conflict data sets and point patterns. In *Modeling Conflict Dynamics with Spatio-Temporal Data. Springer-Briefs in Applied Sciences and Technology* 1–14. Springer, Cham.
- ZEITZOFF, T. (2011). Using social media to measure conflict dynamics: An application to the 2008–2009 Gaza conflict. *J. Confl. Resolut.* 0022002711408014.
- ZHAO, Q., ERDOGDU, M. A., HE, H. Y., RAJARAMAN, A. and SEISMIC, J. L. (2015). A self-exciting point process model for predicting tweet popularity. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1513–1522. ACM, New York.
- ZHENG, P., YUAN, S., WU, X. and WU, Y. (2019). Identifying hidden buyers in darknet markets via Dirichlet Hawkes process. Preprint. Available at [arXiv:1911.04620](https://arxiv.org/abs/1911.04620).
- ZHUANG, J., OGATA, Y. and VERE-JONES, D. (2002). Stochastic declustering of space-time earthquake occurrences. *J. Amer. Statist. Assoc.* **97** 369–380. [MR1941459 https://doi.org/10.1198/016214502760046925](https://doi.org/10.1198/016214502760046925)