DOES TERRORISM TRIGGER ONLINE HATE SPEECH? 
ON THE ASSOCIATION OF EVENTS AND TIME SERIES

BY ERIK SCHARWÄCHTER AND EMMANUEL MÜLLER

University of Bonn, Germany

Hate speech is ubiquitous on the Web, in particular on social media platforms that are driven by user-generated content. Recently, the offline causes that contribute to online hate speech have received increasing attention, but most existing works are based on singular case studies, and there is a lack of longitudinal approaches. A recurring question is whether the occurrence of extreme events offline systematically triggers bursts of hate speech online, indicated by peaks in the volume of hateful social media posts. Formally, this question translates into measuring the association between a time series and a sparse event series. We propose a novel statistical methodology to measure, test and visualize the systematic association between rare exogenous events and peaks in a time series. In contrast to previous methods for causal inference or independence tests on time series, our approach focuses only on the timing of events and peaks, and no other distributional characteristics. We follow the framework of event coincidence analysis (ECA) that was originally developed to correlate point processes. We formulate a discrete-time variant of ECA and derive all the required distributions to enable analyses of peaks in time series, with a special focus on serial dependencies and peaks over multiple thresholds. The analysis gives rise to a novel visualization of the association via quantile-trigger rate plots. We demonstrate the utility of our approach by analyzing whether Islamist terrorist attacks in Western Europe and North America systematically trigger bursts of hate speech and counter-hate speech on Twitter.

1. Introduction. The ubiquity of hate speech in online social media, while distressing, delivers insights into key emotive subjects within a society and globally. Typically, the terms and conditions of online social media platforms prohibit hate speech. Providers ask users to report such contents in order to take further action, e.g., by removing the contents, warning the involved users, or suspending or deleting their profiles (Matias et al., 2015). On the one hand, deletion of harassing material and incitements to violence against individuals is important to protect the victims. On the other hand, bursts of group-based hate speech online, e.g., anti-Muslim, anti-immigrant, anti-black, antisemitic, or homophobic, help to identify triggers and mechanisms of hate and thereby inform policymakers and non-governmental organizations. A recent publication by the United Nations
Educational, Scientific and Cultural Organization (UNESCO) points out that the “character of hate speech online and its relation to offline speech and action are poorly understood” and that the “causes underlying the phenomenon and the dynamics through which certain types of content emerge, diffuse and lead—or not—to actual discrimination, hostility or violence” should be investigated more deeply (Gagliardone et al., 2015).

We propose a novel statistical methodology that enables analyses of the systematic relation between rare offline events and online social media usage. Following a recent study (Olteanu et al., 2018), we demonstrate the utility of our approach by analyzing whether Islamist terrorist attacks systematically trigger bursts of hate speech and counter-hate speech on Twitter. We operationalize these speech acts by tracking usage of the hashtags #stopislam (anti-Muslim hate speech) and #notinmyname (Muslim counter-hate speech), as well as the Arabic keyword kafir (jihadist hate speech against “non-believers”) over a period of three years (2015-2017). We correlate usage of these terms with the occurrence of severe terrorist incidents in Western Europe and North America in the same time period. The key novelty of our approach is that we focus only on the timing of spikes in the resulting time series, not their magnitude or duration as in previous studies on the topic (Olteanu et al., 2018; Burnap et al., 2014). If spikes in the time series coincide with events more often than expected under an independence assumption, there is evidence for a systematic statistical relationship between the two. We thus map the correlation problem into the framework of event coincidence analysis (ECA) (Donges et al., 2016) that was recently proposed to measure coincidences for pairs of (continuous-time) point processes.

For this purpose, we first provide a discrete-time formulation of ECA for pairs of event series that corresponds to the original continuous-time formulation for point processes (Donges et al., 2016). Building on this formulation, our methodological contributions are as follows. We replace the lagging event series with a thresholded time series and derive the null distribution of the ECA statistic for exceedances of a fixed threshold. The derivation is valid for a large class of strictly stationary time series with serial dependencies. Since a single threshold is not sufficient to capture the total association, we further derive the joint distribution of the ECA statistic at multiple thresholds. The analysis gives rise to a surprise measure for a statistical hypothesis test and a novel visualization of trigger coincidences via quantile-trigger rate plots (QTR plots). With our method, we are able to show that Islamist terrorist attacks in Western Europe and North America systematically trigger bursts of anti-Muslim hate speech on Twitter (#stopislam), which confirms case studies and explorative analyses from the literature.
2. Related Work. Most research on the relation between offline actions/events and online hate speech so far is based on case studies. The UNESCO publication mentioned earlier describes a few qualitative case studies on the extreme right-wing online forum Stormfront (De Koster and Houtman, 2008; Bowman-Grieve, 2009; Meddaugh and Kay, 2009), and presents findings from non-academic reports on online hate speech during elections in Kenya and against the Rohingya community in Myanmar (Gagliardone et al., 2015). Burnap et al. (2014) perform a quantitative case study of the social media reaction after the Woolwich terrorist attack in the United Kingdom (May 23rd, 2013). They analyze the size and survival rate of posts that express tension, defined as antagonistic or accusatory content similar to hate speech. In follow-up works (Burnap and Williams, 2014; Williams and Burnap, 2016; Burnap and Williams, 2016), they exploit their findings to train hate speech classifiers and predictive models for information flow following emotive offline events. Magdy, Darwish and Abokhodair (2015) perform a quantitative case study of Twitter usage after the Paris terrorist attacks (November 13th, 2015) and find 21.5% of the posts attacking Islam or Muslims, as opposed to 55.6% defending posts. In contrast to these case studies, we address the systematic relation between offline events and online hate speech in a longitudinal study with 17 relevant events over three years to uncover a potential causal link.

Müller and Schwarz (2019) empirically analyze the relation between online hate speech and hate crimes against refugees in Germany by performing fixed effects panel regression on data that covers the years 2015 to 2017. They exploit internet outages as sources of quasi-experimental exogenous variation to establish a causal link. The major difference to our work is that their events of interest are so numerous that they can be aggregated to a numerical value with weekly resolution and be analyzed with standard statistical methodology. In our setting, events are very rare.

Most related to our work, Olteanu et al. (2018) analyze the impact of Islamist and Islamophobic terrorist attacks on anti-Muslim hate speech online in a longitudinal study with 13 relevant events over 19 months. They perform counterfactual analyses (Brodersen et al., 2015) of a large number of time series representing the daily volumes of hundreds of anti-Muslim keywords independently for every event and report aggregated effects. However, the counterfactual approach is designed for singular, controlled interventions, not for observational studies with reoccurring, uncontrolled events. The reported aggregated effects are thus explorative and not corroborated by measures of statistical significance. We fill this gap by providing a novel statistical methodology to systematically analyze coincidences of rare events and peaks.
in a time series within the framework of ECA (Donges et al., 2016).

ECA was recently developed to measure the association between two types of reoccurring events. It was applied to assess whether floods systematically trigger epidemic outbreaks (Donges et al., 2016), or whether natural disasters systematically trigger violent conflicts (Schleussner et al., 2016). We give an introduction to ECA and discuss challenges when applied to the study of peaks in time series in Section 4. Event synchronization (Quiroga, Kreuz and Grassberger, 2002) is similar to ECA, but allows the time tolerance for coincidences to vary. This increases model expressiveness at the cost of an analytical null distribution. The major difference between our ECA-based approach and other methods for causal inference in time series (Granger, 1969; Box and Tiao, 1975; Schreiber, 2000; Bressler and Seth, 2011; Brodersen et al., 2015) or recent statistical independence tests (Besserve, Logothetis and Schölkopf, 2013; Chwialkowski and Gretton, 2014) is that the only feature it uses is the timing of events and peaks, irrespective of other distributional characteristics. In particular, it does not assume an underlying predictive model that would have to explain the exact behavior of the time series after event occurrences. By focusing on peaks in the time series, it is closely related to measures and models for tail dependence of random variables (Frahm, Junker and Schmidt, 2005; Yan, Wu and Zhang, 2019).

Causal inference techniques have been applied in social media studies before (Cunha, Weber and Pappa, 2017; Chandrasekharan et al., 2017; Saha, Chandrasekharan and De Choudhury, 2019). These works differ from ours in that they do not focus on the association between time series and event series. Recent work on online hate speech has focused on the targets of hate (Silva et al., 2016; Mondal, Silva and Benevenuto, 2017; ElSherief et al., 2018a), characterizations of hateful users (ElSherief et al., 2018a; Ribeiro et al., 2018), as well as geographic (Mondal, Silva and Benevenuto, 2017) and linguistic differences (ElSherief et al., 2018b) in hate. Perhaps the largest body of research on online hate speech in the past decade has been on different approaches for its automatic identification (Warner and Hirschberg, 2012; Kwok and Wang, 2013; Burnap and Williams, 2014; Djuric et al., 2015; Davidson et al., 2017). An overview of the various approaches is given in a recent survey (Schmidt and Wiegand, 2017).

3. Data. For a quantitative analysis of social media usage in reaction to Islamist terrorist attacks we have to operationalize these terms. We only provide a brief overview of the data here and refer to Appendix A for a detailed description and discussion.
# Table 1

Severe Islamist terrorist attacks in Western Europe and North America.

<table>
<thead>
<tr>
<th>Date</th>
<th>City</th>
<th>Date</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-01-07</td>
<td>Paris, France</td>
<td>2016-12-19</td>
<td>Berlin, Germany</td>
</tr>
<tr>
<td>2015-12-02</td>
<td>San Bernardino, USA</td>
<td>2017-04-07</td>
<td>Stockholm, Sweden</td>
</tr>
<tr>
<td>2016-03-22</td>
<td>Brussels, Belgium</td>
<td>2017-05-22</td>
<td>Manchester, UK</td>
</tr>
<tr>
<td>2016-06-12</td>
<td>Orlando, USA</td>
<td>2017-06-03</td>
<td>London, UK</td>
</tr>
<tr>
<td>2016-07-14</td>
<td>Nice, France</td>
<td>2017-08-17</td>
<td>Barcelona, Spain</td>
</tr>
<tr>
<td>2016-07-24</td>
<td>Ansbach, Germany</td>
<td>2017-09-15</td>
<td>London, UK</td>
</tr>
<tr>
<td>2016-09-17</td>
<td>New York City, USA</td>
<td>2017-10-31</td>
<td>New York City, USA</td>
</tr>
<tr>
<td>2016-11-28</td>
<td>Columbus, USA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Islamist terrorist attacks. We obtained a comprehensive list of global terrorist attacks from the publicly available Global Terrorism Database (GTD) (National Consortium for the Study of Terrorism and Responses to Terrorism (START), 2018). For our analysis, we extract the 17 severe Islamist terrorist attacks in Western Europe and North America between January 2015 and December 2017 shown in Table 1.

Social media response. To assess the online social media response to these events in terms of anti-Muslim hate speech, Muslim counter-hate speech, and jihadist hate speech, we retrieved time series of the global Twitter volume in the same time period (2015-2017) for the three keywords #stopislam, #notinmyname and kafir (“non-believer”) that represent the three speech acts. We used the ForSight platform provided by Crimson Hexagon\(^1\) to retrieve daily time series of the global Twitter volume for our keywords. Global daily volume for the queries after preprocessing is shown in Figure 1, along with all Islamist terrorist attacks from Table 1.

4. Methods. Our goal is to analyze the systematic relation between offline events and online social media usage. Formally, we model the occurrence of terrorist attacks by a (discrete-time) event series \(\mathcal{E} = \{E_1, ..., E_T\}\), where each \(E_t\) is a binary random variable with \(E_t = 1\) if and only if there is a terrorist attack at time \(t\), and \(E_t = 0\) otherwise. Social media usage is captured by a (discrete-time) time series \(\mathcal{X} = \{X_1, ..., X_T\}\), where each \(X_t\) is a continuous random variable that indicates the daily volume of posts. A peak in the time series is the exceedance of some large threshold \(\tau \in \mathbb{R}\). The problem is to decide whether the number of events in \(\mathcal{E}\) that trigger peaks in \(\mathcal{X}\) is so high that the association should be considered statistically significant: in this case, there is a potential causal link between event occurrences and

\(^1\)https://www.crimsonhexagon.com/
peaks in the time series. Observe that for a time series $X$ and threshold $\tau$, the threshold exceedance series

$$A = \{I(X_1 > \tau), ..., I(X_T > \tau)\}$$

is itself an event series. Here, $I(C)$ is an indicator function that is 1 if and only if the condition $C$ is true, and 0 otherwise (c.f. Figure 2). The threshold exceedance series contains only information on the timing of exceedances, and disregards all other distributional characteristics. The problem of correlating event series with peaks in a time series can thus directly be mapped to the problem of correlating two event series, e.g., using ECA.

**Challenges.** There are, however, two challenges that need to be addressed when applying measures designed for pairs of event series in this context:
serial dependencies and threshold selection. Serial dependencies in the time series lead to clustering of events in the threshold exceedance series. This effect can be observed when comparing the upper and the lower time series in Figure 2. A potential correlation measure must correctly handle this clustering of events in the assessment of statistical significance. The second challenge is that the choice of threshold has a strong impact on the results of the analysis, but is often not straight-forward. In fact, the magnitude of the peak may vary from event to event: a full picture of the association between events and peaks can only be obtained when considering exceedances at multiple thresholds at the same time.

We now proceed with a detailed exposition of the statistical methodology that we propose for the analysis. We embed our contributions within the existing framework of ECA. We begin with a discrete-time formulation of ECA for pairs of event series in Section 4.1 that corresponds to the original continuous-time formulation for point processes (Donges et al., 2016). In Section 4.2 we address the challenge of serial dependencies by deriving a novel analytical null distribution for the ECA statistic that is valid for threshold exceedance series for a large class of strictly stationary time series. We then derive the joint null distribution for coincidences at multiple thresholds in Section 4.3 and apply a natural surprise measure to assess statistical significance. We complement the analytical results with a novel visualization of the association via quantile-trigger rate (QTR) plots.

4.1. Discrete-time event coincidence analysis. ECA is a statistical methodology to assess whether two types of events are independent or whether one kind of event systematically triggers or precedes the other kind of event. The basic idea of ECA is to count how many times the two kinds of events coincide, and assess whether this number is statistically significant under the assumption of independence.

4.1.1. Definition. Let $A = \{E_A^1, \ldots, E_A^T\}$ and $B = \{E_B^1, \ldots, E_B^T\}$ be two event series of length $T$ with $E_A^t, E_B^t \in \{0, 1\}$ for all $t$, and $N_A = \sum_t E_A^t$ and $N_B = \sum_t E_B^t$ event occurrences. Furthermore, let $\Delta \in \mathbb{N}_0$ be a user-defined time tolerance. ECA measures the extent to which $B$ events precede $A$ events, with a time tolerance of $\Delta$. We thus refer to $A$ as the lagging and $B$ the leading event series. ECA considers two possibilities to measure this extent: trigger coincidences and precursor coincidences. A trigger coincidence occurs whenever a $B$ event triggers an $A$ event within the next $\Delta$ time steps, whereas a precursor coincidence occurs whenever an $A$ event is preceded by a $B$ event within the previous $\Delta$ time steps. The two types of coincidences are illustrated in Figure 3. In the example, there are three
trigger coincidences and four precursor coincidences. For $\Delta = 0$ they are identical, but for $\Delta > 0$ they differ, in general. A significant number of trigger or precursor coincidences indicates a possible causal link from $\mathcal{B}$ to $\mathcal{A}$. The opposite direction can be analyzed analogously by exchanging the labels. Formally, the number of trigger coincidences is defined as

$$K_{tr} = K_{tr}^\Delta(\mathcal{B}, \mathcal{A}) := \sum_{t=1}^{T-\Delta} E^B_t \cdot \left( \max_{\delta=0,\ldots,\Delta} E^A_{t+\delta} \right)$$

and the number of precursor coincidences as

$$K_{pre} = K_{pre}^\Delta(\mathcal{B}, \mathcal{A}) := \sum_{t=\Delta+1}^{T} E^A_t \cdot \left( \max_{\delta=0,\ldots,\Delta} E^B_{t-\delta} \right).$$

The order of the function arguments $\mathcal{B}$ and $\mathcal{A}$ corresponds to the temporal ordering that is analyzed (and thus the potential causal direction). We omit the parameter $\Delta$ and the function arguments whenever they are clear from the context. The corresponding coincidence rates are given by $r_{tr} := K_{tr}/N_B$ and $r_{pre} := K_{pre}/N_A$. In the example from Figure 3, we have $r_{tr} = 1$ and $r_{pre} = \frac{2}{3}$. A high trigger coincidence rate indicates that a large fraction of $\mathcal{B}$ events is followed by an $\mathcal{A}$ event. In other words, $\mathcal{B}$ events systematically trigger $\mathcal{A}$ events. A high precursor coincidence rate indicates that a large fraction of $\mathcal{A}$ events is preceded by a $\mathcal{B}$ event, i.e., the occurrence of $\mathcal{A}$ events can be explained (almost) exclusively by $\mathcal{B}$ events. The two measures are complementary and should be selected based on the research question.

4.1.2. Null distribution. To assess whether an observed trigger coincidence rate is statistically significant, we need the null distribution of $K_{tr}$ under the assumption that the processes are independent. For this purpose, we introduce the binary random variables $Z^A_t := \max_{\delta=0,\ldots,\Delta} E^A_{t+\delta}$ for all $t = 1, \ldots, T - \Delta$. 

\[ \text{Fig 3. Trigger coincidences and precursor coincidences for two event series } \mathcal{A} \text{ and } \mathcal{B}, \text{ with time tolerance } \Delta = 4. \]
that indicate whether there is an \( A \) event in the window \( t, ..., t + \Delta \). This allows rewriting Equation 2 as

\[
K_{tr} = \sum_{t=1}^{T-\Delta} E_t^B \cdot Z_t^A = \sum_{t; E_t^B = 1} Z_t^A
\]

and reveals that the number of trigger coincidences is effectively a sum of Bernoulli trials, each associated with an event occurrence in \( B \). Sums over a fixed number of independent and identically distributed (iid) Bernoulli trials follow binomial distributions. However, for an arbitrary event series \( A \), \( Z_t^A \) and \( Z_t^A' \) may be neither identically distributed nor independent. Additional assumptions are required to derive the null distribution analytically.

In a simple and analytically tractable case, \( A \) and \( B \) are independent Bernoulli processes, i.e., \( A \perp \perp B, E_1^A, ..., E_T^A \sim \text{Bernoulli}(p_A) \) and \( E_1^B, ..., E_T^B \sim \text{Bernoulli}(p_B) \), with \( P(E_t^A = 1) = p_A \) and \( P(E_t^B = 1) = p_B \). Then all \( Z_t^A \) for \( t = 1, ..., T - \Delta \) are identically distributed with success probability

\[
P(Z_t^A = 1) = 1 - P(E_t^A = 0, ..., E_{t+\Delta}^A = 0)
\]

\[
= 1 - (1 - p_A)^{\Delta+1}.
\]

Furthermore, \( Z_t^A \) and \( Z_t^A' \) are independent whenever they are separated by more than \( \Delta \) time steps. Under the additional assumption that the \( N_B \) events in \( B \) are separated by more than \( \Delta \) time steps, the conditional null distribution of \( K_{tr} \) is thus

\[
K_{tr} | N_B \sim \text{Binomial} \left( N_B, 1 - (1 - p_A)^{\Delta+1} \right).
\]

Since for a Bernoulli process \( B \), \( N_B \) is itself binomially distributed, the unconditional null distribution is

\[
K_{tr} \sim \text{Binomial} \left( T, p_B \cdot (1 - (1 - p_A)^{\Delta+1}) \right).
\]

4.1.3. Statistical test procedure. Following the framework of ECA, we use \( K_{tr} \) as a test statistic to decide between the null hypothesis of independence of \( A \) and \( B \), and the alternative hypothesis of a trigger relationship. If the number of trigger coincidences is unusually large, the null hypothesis is rejected in favor of the alternative hypothesis. For any observed pair of series, the numbers of \( A \) and \( B \) events are fixed, such that the conditional null distribution for \( K_{tr} \) from Equation 7 must be used. The success probabilities are estimated as \( \hat{p}_A = \frac{N_A}{T} \) and \( \hat{p}_B = \frac{N_B}{T} \), respectively. The \( p \)-value for an
observed number of trigger coincidences $k_{tr}$ is obtained from the probability mass function of the binomial distribution in Equation 7,

$$(9) \quad P(K_{tr} \geq k_{tr} \mid N_B) = \sum_{k=k_{tr}}^{N_B} \binom{N_B}{k} \cdot \pi^k \cdot (1-\pi)^{N_B-k},$$

where $\pi = 1 - (1 - \hat{p}_A)^{\Delta+1}$. The null hypothesis is rejected at the desired significance level $\alpha$ if $P(K_{tr} \geq k_{tr} \mid N_B) < \alpha$. The null distribution and test for significance for the number of precursor coincidences can be derived completely analogously. These results are valid for event series that follow Bernoulli processes; they correspond to the null distributions derived for homogeneous Poisson processes in continuous-time ECA (Donges et al., 2016). For other processes, such analytical results have not been obtained, and to date, Monte Carlo methods are required to simulate the null distribution.

4.2. Coincidences with threshold exceedances. Threshold exceedance series obtained from time series by applying fixed thresholds as in Equation 1 are often clustered: they do not, in general, follow a Bernoulli process, and the results from the previous section therefore do not apply. We now explicitly focus on the null distribution of the number of trigger coincidences when the lagging event series $A$ is a threshold exceedance series. Our key observation is that in this case $K_{tr}$ can be reformulated such that the Extremal Types Theorem from Extreme Value Theory (Coles, 2001) is applicable. We first define the number of trigger coincidences for a leading event series $E$, lagging time series $X$, time tolerance $\Delta \in \mathbb{N}_0$, and a fixed threshold $\tau \in \mathbb{R}$ by plugging the threshold exceedance series into Equation 2:

$$(10) \quad K_{tr} = K_{tr}^{\Delta,\tau}(E, X) := \sum_{t=1}^{T-\Delta} E_t \cdot \left( \max_{\delta=0,\ldots,\Delta} I(X_{t+\delta} > \tau) \right).$$

As before, the order of the function arguments indicates the lead-lag relation (and potential causal direction), and we omit parameters and arguments if clear from the context. Observe that we can swap the order of the max-operator and the indicator function

$$(11) \quad \max_{\delta=0,\ldots,\Delta} I(X_{t+\delta} > \tau) = I\left( \max_{\delta=0,\ldots,\Delta} X_{t+\delta} > \tau \right).$$

For brevity of notation, we introduce the helper variables

$$(12) \quad Z_t^{\Delta,\tau} := I\left( \max_{\delta=0,\ldots,\Delta} X_{t+\delta} > \tau \right).$$
such that

$$K_{tr} = \sum_{t=1}^{T-\Delta} E_t \cdot Z_t^{\Delta, \tau} = \sum_{t:E_t=1} Z_t^{\Delta, \tau}. \quad (13)$$

The number of trigger coincidences is again a sum of Bernoulli trials, each associated with one of the $N_E$ event occurrences in $E$. If the time series $X$ is strictly stationary, the Bernoulli trials are all identically distributed with the same marginal distribution $P(Z_t^{\Delta, \tau})$, but they are, in general, not independent.

The benefit of swapping the max-operator and the indicator function is that now the success probability of the Bernoulli trials

$$P(Z_t^{\Delta, \tau} = 1) = P\left(\max_{\delta=0,\ldots,\Delta} X_{t+\delta} > \tau\right) \quad (14)$$

$$= 1 - P\left(\max_{\delta=0,\ldots,\Delta} X_{t+\delta} \leq \tau\right) \quad (15)$$

is defined directly on the time series, not on the event series as in Equation 5. Probabilities in the form of Equation 15 have been studied extensively in Extreme Value Theory. In fact, according to the Extremal Types Theorem (ETT), the maximum of a large number of iid random variables is approximated by the Generalized Extreme Value (GEV) distribution, under mild constraints on the underlying distribution of the random variables:

**Theorem 1 (Extremal Types Theorem (Coles, 2001)).** Let $X_1, \ldots, X_n \sim F$ and $M_n = \max_{i=1,\ldots,n} X_i$. If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \to G(z) \text{ as } n \to \infty$$

for a non-degenerate distribution function $G$, then $G$ is a member of the GEV family

$$G(z) = \exp\left\{ -\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}} \right\},$$

defined on $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$.

The ETT was shown to also apply more generally to the maxima of strictly stationary time series, as long as they fulfill an additional property that
eliminates long-range dependencies; see Coles (2001, ch. 5.2) for technical details. Thus, for strictly stationary $\mathcal{X}$ with limited long-range dependencies, and for large $\Delta$, the variables $Z_{t}^{\Delta, \tau}$ are identically distributed with

$$P(Z_{t}^{\Delta, \tau} = 1) \approx 1 - G(\tau; \theta_{\Delta}) \quad (16)$$

that depends on the threshold $\tau$ and the time tolerance $\Delta$. The normalizing constants from the ETT disappear in the GEV parameter vector $\theta_{\Delta} = (\xi, \mu, \sigma)$. The subscript in the parameter vector denotes its dependence on $\Delta$. The larger $\Delta$ and the higher $\tau$, the better the approximation by the GEV distribution.\(^2\) If there are only limited long-range dependencies in the time series, the variables $Z_{t}^{\Delta, \tau}$ associated with the $N_{E}$ event occurrences in $\mathcal{E}$ can further be viewed as approximately independent. Under the above assumptions, the null distribution of the number of trigger coincidences for $N_{E}$ events is approximated by the binomial

$$K_{tr}^{\tau, \Delta} \mid N_{E} \sim \text{Binomial}(N_{E}, 1 - G(\tau; \theta_{\Delta})). \quad (17)$$

If the leading event series $\mathcal{E}$ is a Bernoulli process with success probability $p_{E}$, the unconditional null distribution is approximated by the binomial

$$K_{tr}^{\tau, \Delta} \sim \text{Binomial}(T, p_{E} \cdot (1 - G(\tau; \theta_{\Delta}))). \quad (18)$$

Regardless of the ultimate choice of the threshold $\tau$, the parameters $\theta_{\Delta}$ can be estimated by splitting the time series $\mathcal{X}$ into consecutive blocks of size $\Delta + 1$ and fitting the GEV distribution to the maxima of each block, e.g., using maximum likelihood. After estimating the parameters on $\mathcal{X}$ and applying the threshold $\tau$, the statistical significance of the observed number of trigger coincidences $k_{tr}$ can be assessed using the $p$-value from Equation 9 with the revised estimate $\pi = 1 - G(\tau; \theta_{\Delta})$ as in Equation 17.

**Examples.** With these results, we are able to apply discrete-time ECA to analyze the triggers of exceedances of a fixed threshold. Fixed thresholds may be derived from domain-specific hypotheses. For example, we might want to test whether Islamist terrorist attacks systematically trigger bursts of more than 1,000 posts per day on Twitter that contain the hashtag #stopislam. We observe $K_{tr} = 9$ trigger coincidences within a time tolerance of $\Delta = 7$ days after the events, which gives a trigger coincidence rate of $r_{tr} = .53$. Assuming that the raw Twitter time series is strictly stationary, we obtain a $p$-value of $p = .0014$ using our GEV-based null distribution. Lacking a

\(2\text{In Section B.1, we demonstrate that a value of } \Delta = 7 \text{ is already large enough for the approximation to be valid.}\)
specific hypothesis for the value of the threshold, generic values can be tested, such as the empirical 95%-quantile or 99%-quantile. For example, we can test the hypothesis that Islamist terrorist attacks systematically trigger bursts of #notinmyname usage that exceed the volume of 95% of all days.

We observe $K_{tr} = 4$ trigger coincidences with $\Delta = 7$, which gives a trigger coincidence rate of $r_{tr} = .24$. Again assuming that the raw Twitter time series is strictly stationary, we obtain a $p$-value of $p = .2427$. These examples are only illustrative, since the raw Twitter time series are not strictly stationary. In Appendix A, we describe the preprocessing scheme that we use for the results in Section 5 to make the time series stationary.

4.3. Exceedances of increasing thresholds. In case a fixed threshold is unavailable, or if a full picture of the association with peaks of various magnitudes is required, exceedances at multiple thresholds have to be considered. In fact, threshold exceedances at multiple levels are highly dependent: if an observation exceeds any threshold $\tau$, it also exceeds all lower thresholds. The numbers of trigger coincidences at multiple thresholds are thus dependent as well. To enable joint analyses of multiple threshold exceedances and thereby eliminate the need of selecting a single fixed threshold, we now derive the joint null distribution of trigger coincidences at multiple thresholds, along with a novel visualization for this statistical association.

4.3.1. Trigger coincidence processes. Let $\tau = (\tau_1, ..., \tau_M)$ be a sequence of increasing thresholds $\tau_1 < ... < \tau_M$. The trigger coincidence process

$$K_{tr} = K_{\Delta \tau}^E (\mathcal{E}, \mathcal{X}) = \left( K_{tr}^{\Delta \tau_1}, ..., K_{tr}^{\Delta \tau_M} \right)$$

is the corresponding sequence of the numbers of trigger coincidences for all given thresholds $\tau$. The trigger coincidence process is monotonically decreasing with increasing thresholds, and the number of exceedances at level $i$ depends on the number of exceedances before $i$. The canonical trigger coincidence process is given by the threshold sequence $\tau = (\tau_1, ..., \tau_M) = (X_{(1)}, ..., X_{(T)})$, where $X_{(t)}$ denotes the order statistic of the time series such that $X_{(1)} < ... < X_{(T)}$. Trigger coincidence processes for other sequences of thresholds in the range $[X_{(1)}, X_{(T)}]$ approximate the canonical trigger coincidence process. Figure 4 illustrates the concept in a simulated example, details on the simulation are given below.

Example. We generated a time series of length $T = 4096$ from iid standard exponential random variables, applied a moving average (MA) filter of order 8, standardized the result and subtracted the minimum value to obtain a non-negative time series $\mathcal{X}$ with serial dependencies. We then generated
two event series: an independent and a dependent one. To simulate a peak
trigger relationship, we randomly sampled $N_E = 32$ time steps $t$ from the
time series where $X_t > 4$, and set $E_{t-4} = 1$ for these $t$ in the dependent
event series. In the independent event series, we distributed the 32 events
completely at random. An excerpt of the resulting series is shown in the
upper part of Figure 4. The two canonical trigger coincidence processes of
length $M = T$ for $\Delta = 7$ are depicted in the lower left part of Figure 4. At
low thresholds, large numbers of trigger coincidences are observed both for
the dependent and the independent event series. For higher thresholds, the
numbers of trigger coincidences for the dependent event series dramatically
exceed the numbers of the independent event series. By construction, all 32
events in the dependent series trigger an exceedance of the threshold 4; see
the marker (\textAst). The threshold 5 is exceeded after 13 out of 32 events from
the dependent event series, while it is only exceeded after a single event from
the independent event series; see the marker (\textDagger).

4.3.2. Quantile-trigger rate plots. Plots of trigger coincidence processes
as in Figure 4 (bottom left) help to visually assess whether events in $E$
systematically trigger peaks of various magnitudes in a time series $X$. However,
the scales of the axes depend on the range of values in $X$ and the number of
events $N_E$, which makes it hard to compare these plots across multiple pairs
of time series and event series. Furthermore, the absolute threshold value

Fig 4. Canonical trigger coincidence processes (left) and corresponding QTR plot (right),
for a simulated time series and two event series (independent and dependent).
is not informative about the actual extremeness of a peak with respect to the bulk of the data. Therefore, we propose **quantile-trigger rate (QTR) plots** as a standardized visualization of trigger coincidence processes with normalized axes. In a QTR plot, the $x$-axis is normalized by using empirical $p$-quantiles from $X$ instead of the absolute thresholds $\tau_m$, while the $y$-axis is normalized by using the trigger coincidence rate $r_{tr}$ instead of the absolute number of trigger coincidences $K_{tr}$. The QTR plot for the example above is shown in Figure 4 (bottom right). The most striking difference is that now the dependent curve appears much more extreme, since the thresholds larger than 4 correspond to very high empirical $p$-quantiles. Intuitively, the closer an observed trigger coincidence process to the top-right corner of the QTR plot, the more events coincide with threshold exceedances, at more extreme levels. To obtain a detailed picture of the association at higher $p$-quantiles, the $x$-axis in a QTR plot can be cropped accordingly.

However, QTR plots have to be interpreted with care. The shape of a trigger coincidence process for an *independent* pair of event series and time series in a QTR plot depends on the statistical properties of the input data. Let $X$ be an iid time series, $E$ be an independent Bernoulli process, and $\Delta = 0$. In this case, the fraction of events that coincide with an exceedance of the empirical $p$-quantile is exactly $1 - p$. The trigger coincidence process is a straight line from $(0, 1)$ to $(1, 0)$ in a QTR plot, regardless of the other characteristics of $X$ and $E$. Figure 5 illustrates the impact of serial dependencies in $X$ and the time tolerance $\Delta$ on the shape of the trigger coincidence process under independence in a QTR plot. With increasing time tolerance $\Delta$, there are more trigger coincidences under independence, and the lines in the QTR plot move towards the top-right corner of the plot. This effect is strongest for iid time series, but also occurs for time series with serial dependencies. Thus, a line that bends towards the top-right corner of the QTR plot is necessary, but not sufficient to conclude a trigger relationship. We need a statistical test that operates on the trigger coincidence process to assess whether the shape in a QTR plot is unusual under an independence assumption. For this purpose, we now derive the statistical properties of the trigger coincidence process $K_{tr}^{\Delta, \tau}$ for multiple thresholds $\tau = (\tau_1, ..., \tau_M)$ with $\tau_1 < ... < \tau_M$.

4.3.3. **Markov chain model.** The main result in Section 4.2 is that the number of trigger coincidences $K_{tr}^{\Delta, \tau}$ for a *single* threshold $\tau$ follows a GEV-based binomial distribution. This derivation allows only pointwise rejections of independence for the respective threshold. Given the strong dependencies between coincidences at multiple thresholds, pointwise decisions provide only a limited picture of the association. To assess whether a *complete* trigger
Fig 5. Expected QTR plots for three time series with different levels of serial dependencies (MA orders 0, 32, 128) and independent event series. For every MA order, we simulate a single time series of length $T = 4096$ from the exponential moving average model described above, and select 50 thresholds at equally spaced $p$-quantiles between 0 and 1. For every threshold $\tau$ and every $\Delta \in \{0, 1, 2, 4, 8, 16, 32, 64\}$, we estimate the expected trigger coincidence rate $r_{\Delta, \tau} = K_{\Delta, \tau}/N_E$ for an independent event series by simulating 100 independent event series with $N_E = 32$ events and averaging the trigger coincidence rates over the 100 runs. Note that for large $\Delta$ and $\tau$, this expectation can be approximated by the expected value of our GEV-based binomial distribution from Equation 17.

The coincidence process is so unusual that the null hypothesis of independence has to be rejected, we need the joint distribution $P(K_{\Delta, \tau}^{\Delta, \tau_1}, ..., K_{\Delta, \tau}^{\Delta, \tau_M})$. In general, $P(K_{\Delta, \tau}^{\Delta, \tau_1}, ..., K_{\Delta, \tau}^{\Delta, \tau_M})$ can be decomposed as

$$P(K_{\Delta, \tau}^{\Delta, \tau_1}) \cdot \prod_{i=2}^{M} P(K_{\Delta, \tau}^{\Delta, \tau_i} | K_{\Delta, \tau}^{\Delta, \tau_1}, ..., K_{\Delta, \tau}^{\Delta, \tau_{i-1}}).$$

We have already derived the marginal distribution $P(K_{\Delta, \tau}^{\Delta, \tau_1})$ and now focus on the conditionals. Suppose there is an exceedance of the threshold $\tau_{i-1}$ in $X$ within the window $t,...,t+\Delta$, i.e., $Z_{\Delta, \tau_1}^{\Delta, \tau_1} = 1$. To derive the probability that there is also an exceedance of the higher threshold $\tau_i$, observe that

$$P(Z_{\Delta, \tau_i}^{\Delta, \tau_1} = 1 | Z_{\Delta, \tau_1}^{\Delta, \tau_1} = 1) = \frac{P(Z_{\Delta, \tau_i}^{\Delta, \tau_1} = 1)}{P(Z_{\Delta, \tau_1}^{\Delta, \tau_1} = 1)} \approx \frac{1 - G(\tau_i; \theta_\Delta)}{1 - G(\tau_{i-1}; \theta_\Delta)},$$

where we used Equation 16 for the approximation. Equation 21 is valid whenever the marginal approximation by the GEV is admissible. Observations in the time series that do not exceed the threshold $\tau_{i-1}$ cannot exceed the higher threshold $\tau_i$, i.e., $P(Z_{\Delta, \tau_i}^{\Delta, \tau_1} = 1 | Z_{\Delta, \tau_1}^{\Delta, \tau_1} = 0) = 0$. With these two results we can specify the conditional distribution for the number of trigger coincidences at level $i$, given the number of coincidences at level $i-1$. We
rewrite the conditional random variable $K_{tr}^{\Delta,\tau_i-1} \mid K_{tr}^{\Delta,\tau_i}$ by restricting the summation in Equation 13 to time steps with $E_t = 1$ and $Z_t^{\Delta,\tau_i-1} = 1$,

$$K_{tr}^{\Delta,\tau_i} \mid K_{tr}^{\Delta,\tau_i-1} = \sum_{t:E_t = 1, Z_t^{\Delta,\tau_i-1} = 1} Z_t^{\Delta,\tau_i}. \quad (22)$$

Under the null hypothesis of independence of $\mathcal{E}$ and $\mathcal{X}$, this is the sum of $K_{tr}^{\Delta,\tau_i-1}$ identically distributed Bernoulli trials with success probability given by Equation 21. Under the absence of long-range dependencies, the individual variables $Z_t^{\Delta,\tau_i}$ are approximately independent, and the conditional number of trigger coincidences follows the binomial distribution

$$K_{tr}^{\tau_i,\Delta} \mid K_{tr}^{\tau_i-1,\Delta} \sim \text{Binomial} \left( K_{tr}^{\tau_i-1,\Delta}, \frac{1 - G(\tau_i; \theta_\Delta)}{1 - G(\tau_{i-1}; \theta_\Delta)} \right). \quad (23)$$

We thus model the conditional distributions as a Markov chain $\mathbb{P}(K_{tr}^{\Delta,\tau_i} \mid K_{tr}^{\Delta,\tau_i}, ..., K_{tr}^{\Delta,\tau_i-1}) = \mathbb{P}(K_{tr}^{\Delta,\tau_i} \mid K_{tr}^{\Delta,\tau_i-1})$. The null distribution of the trigger coincidence process $\mathcal{K}_{tr}$ under independence is fully described by Equation 17 for the smallest threshold and Equation 23 for all larger thresholds.

4.3.4. Surprising trigger coincidence processes. With a full model for the null distribution of the trigger coincidence process $\mathcal{K}_{tr}$ we can compute the likelihood $\mathbb{P}(\mathcal{K}_{tr})$ of any observed sequence of trigger coincidences. If we observe a coincidence sequence that is very unlikely under the null hypothesis, we have evidence that events systematically trigger peaks of various magnitudes in the time series. The trigger coincidence process is a high-dimensional multivariate discrete random variable, where every single realization—even the mode of the distribution—has a very small likelihood. For this reason, we need a way to quantify whether the observed likelihood is unusually small with respect to the distribution of the likelihood values under independence. Equivalently, we can consider the negative log-likelihood $S(\mathcal{K}_{tr}) = -\log \mathbb{P}(\mathcal{K}_{tr})$ as a surprise measure, which has better numerical properties than directly using the likelihood. We then need a way to quantify whether the observed surprise value is unusually large with respect to the distribution of the surprise value under independence. We thus use $S$ as a test statistic in a statistical hypothesis testing framework and reject the null hypothesis of independence at the significance level $\alpha$ if the $p$-value $\mathbb{P}(S \geq s) < \alpha$, where $s = S(\mathcal{K}_{tr}(\mathcal{E}, \mathcal{X}))$ is the surprise value of the observed trigger coincidence process. Lacking an axiomatic derivation for the null distribution of $S$ under independence, we follow a Monte Carlo approach to approximate the $p$-value. We generate $R$ independent event
series $E'$ with the same number of events as $E$ by randomly permuting $E$. For each independent event series, we determine the simulated surprise value $s' = S(K_{tr}(E', X))$. The Monte Carlo $p$-value (Davison and Hinkley, 1997) is given by $\hat{p} = \frac{1 + |\{s'| s'| s\rangle|}{R+1}$. We reject the null hypothesis of independence at significance level $\alpha$ in favor of the alternative hypothesis that events trigger peaks of various magnitudes if $\hat{p} < \alpha$.

5. Results and Discussion. We now utilize our statistical methodology to analyze whether severe Islamist terrorist attacks in Western Europe and North America systematically trigger bursts of hate speech or counter-hate speech (\#stopislam, \#notinmyname, and \#kafer) on Twitter.

Setup. We use the data described in Section 3, which spans the time period from January 1st, 2015 to December 31st, 2017, i.e., a total of 1,096 days with 17 events. We choose a time tolerance of $\Delta = 7$ days to allow enough time for the news about the incidents to spread globally. The simulation study in Section B.1 shows that this time tolerance is also large enough for our GEV-based null distributions to be accurate. For every social media time series $X_i$, we estimate a GEV distribution $G_i$ by splitting $X_i$ into consecutive blocks of size $\Delta + 1$ and fitting the parameters of $G_i$ to the block maxima by maximum likelihood estimation. We then select $M = 32$ thresholds $\tau_i = (\tau_{i,1}, ..., \tau_{i,32})$ at equidistant $p$-quantiles between .75 and 1 from $X_i$, and use the GEV distribution $G_i$ to obtain the parameters of the Markov chain model $P(K_{tr}^{\Delta \tau_i}(E, X_i))$ described by Equation 17 and Equation 23 for independent event series $E$ with $N_E = 17$ events. We compute the observed trigger coincidence processes between the terrorist attack event series $E$ and all social media time series $X_i$, and obtain their surprise values from the respective Markov chain models. We assess the significance of the surprise values with $R = 10,000$ Monte Carlo simulations.

Results. QTR plots with Monte Carlo $p$-values obtained from the statistical test are depicted in Figure 6. The plots are augmented with the marginally expected trigger coincidence processes under independence and the marginal 95% confidence intervals to additionally assess pointwise exceedances of fixed thresholds. The analysis shows that Islamist terrorist attacks in Western Europe and North America systematically trigger bursts of anti-Muslim hate speech on Twitter (\#stopislam, $\hat{p} = .0317$). 90% of Islamist terrorist attacks triggered an exceedance of the .85-quantile, and 60% of Islamist terrorist attacks even triggered an exceedance of the .95-quantile. Our results confirm the findings of previous quantitative studies Burnap et al. (2014); Magdy, Darwish and Abokhodair (2015); Olteanu et al. (2018) with a novel statistical methodology and a larger study period: there is a clear systematic relationship
Fig 6. QTR plots for severe Islamist terrorist attacks and their online social media response.
between Islamist extremist violence offline and anti-Muslim hate speech online.

On the other hand, our analysis does not provide evidence for a systematic association between Islamist terrorist attacks and peaks in jihadist hate speech (kafir, \( \hat{p} = .2075 \)) or Muslim counter-hate speech (\#notinmyname, \( \hat{p} = .3561 \)) in the study period. We stress that individual terrorist attacks may still have triggered such a social media response. Visual inspection of the data in Figure 1 suggests peaks in the hashtag \#notinmyname for Islamist terrorist attacks before July 2016. Hashtag usage is typically subject to trends, so a systematic relationship can only be established for hashtags that are used consistently throughout the study period. The impact of an individual terrorist attack on the social media time series can be assessed, e.g., with counterfactual analyses (Brodersen et al., 2015), regardless of trends. Figure 6 shows that even for jihadist hate speech and Muslim counter-hate speech, the observed numbers of trigger coincidences fall outside the pointwise 95% confidence intervals for some thresholds. Pointwise tests at these specific thresholds as described in Section 4.2 would reject the null hypothesis of independence in favor of a systematic trigger relationship. Our approach of using multiple thresholds within a single hypothesis test thus avoids the dangers of data dredging, which would otherwise require adjustments for multiple hypothesis testing. To validate the results from our multiple threshold test, we computed all \( p \)-values for the pointwise tests at all thresholds and used different adjustment methods that control the family-wise error rate at level \( \alpha = .05 \): Bonferroni, single-step Šidák, step-down Holm in its original variant and in the Šidák variant (Dudoit and van der Laan, 2007). For all multiple test adjustment methods, the results agree with our multiple threshold approach.

Sensitivity analysis. To assess the stability of the results, we further experimented with different choices for the time tolerance \( \Delta = 4...16 \) (ceteris paribus). We found that for \( \Delta = 4...8 \), the results of all tests on all time series are unchanged. For \( \Delta = 9...14 \) our multiple threshold test fails to reject the null hypothesis for the \#stopislam time series, while the multiple pointwise test procedures still reject. For \( \Delta = 15 \) only the Šidák procedures reject the null hypothesis on \#stopislam, while for \( \Delta = 16 \) no test procedure rejects the null hypotheses on any time series. Choosing a time tolerance \( \Delta \) that is longer than necessary thus reduces the sensitivity of the tests. We also varied the number of thresholds \( M \) between 8 and 64 (ceteris paribus), which did not change the outcome any test. At last, we moved the thresholds upwards to more extreme levels by choosing equidistant \( p \)-quantiles from the ranges \(.85 \) to \(1\) and \(.95 \) to \(1\), respectively (ceteris paribus). The outcomes on the \#stopislam and kafir time series remain unchanged, while our multiple
threshold test now detects an additional trigger relationship for \#notinmyname that is not detected by the multiple pointwise test procedures. Overall, the trigger relationship for \#stopislam is very stable across all test procedures with different parameterizations, whereas the results on \#notinmyname are inconclusive.

6. Conclusion. We have refined the statistical methodology to infer potential causal links between an event series and peaks in a time series. Rooted in the framework of event coincidence analysis, we focus only the timing of events and peaks, and no other distributional characteristics. We have derived analytical expressions for the null distributions of the ECA statistic for coincidences with exceedances of a single threshold and multiple thresholds, even under serial dependencies in the time series. Our results are valid for all strictly stationary time series with limited long-range dependencies. They require individual event occurrences to be separated by a sufficient number of time steps such that the GEV-based binomial null distributions well approximate the actual null distributions. Our analysis is therefore most suitable for sparse event series, e.g., representing the occurrences of extreme events. We have made no other assumptions on the time series or the event series that restrict the applicability of our results. Therefore, the ECA-based approach is highly versatile and enables the discovery of unknown associations in a large variety of fields. For a complete causal analysis, confounding factors that influence both series must still be ruled out. This is a challenging direction for future work, as it either requires the specification of a joint model for the confounding factors, the time series, and the event series in the spirit of Granger causality (Granger, 1969), or non-trivial changes in the nonparametric procedure of ECA. A first step in this direction is conditional ECA and joint ECA (Siegmund et al., 2016). We support the methodological contributions described in this work with a simulation study in Appendix B.

SUPPLEMENTARY MATERIAL

Supplement A: Source codes and data
(doi: xyz; supplA.zip). Codes for the simulations that reproduce the Figures 2, 4, 5, 7–9, as well as codes and data to reproduce all results on the hate speech problem (Table 1, Figures 1 and 6).

APPENDIX A: DETAILED DATA DESCRIPTION

Islamist terrorist attacks. We filtered the Global Terrorism Database (GTD) (National Consortium for the Study of Terrorism and Responses to Terrorism (START), 2018) for attacks that hit countries in Western Europe and North
America between January 2015 and December 2017, left at least 10 people wounded, and were conducted by the so-called *Islamic State of Iraq and the Levant* (ISIL), *Al-Qaida in the Arabian Peninsula* (AQAP), Jihadi-inspired extremists or Muslim extremists, according to the GTD.

**Social media response.** The hashtag #stopislam has been observed in anti-Muslim hate speech before (Magdy, Darwish and Abokhodair, 2015; Olteanu et al., 2018) and has also received some media attention (Dewey, 2016; Hemmings, 2016). Many posts that contain the hashtag actually condemn its usage, so spikes in the volume should not be seen as pure bursts of hate speech. Yet, such condemnation is typically triggered by initial anti-Muslim posts. Due to the mixed usage, the magnitude of a spike is no indicator for the extent of online hate, only the presence of a spike is informative.

The phrase “not in my name” is used by members of a group to express their disapproval of actions that are associated with that group or (perceived or actual) representatives of the group (Tormey, 2006; Čehajić and Brown, 2008). It was observed, for example, during global protests against the 2003 war of the US-led coalition against Iraq (Bennett, 2005), or more recently during protests sparked by the murder of a Muslim boy by Hindu nationalists in India 2017 (Krishnan, 2017). Most importantly for the present study, Muslim social media users have repeatedly used the hashtag after Islamist terrorist attacks (Davidson, 2014). Due to the generic nature of the phrase, it cannot solely be viewed as Muslim counter-hate speech. Nonetheless, online social media posts that contain #notinmyname right after Islamist terrorist attacks are likely to convey a Muslim counter-hate message.

At last, the Arabic word *kafir* translates to the English word “non-believer.” It is traditionally used by Muslim fundamentalists against other Muslims that do not adhere to the fundamentalist ideology (Alvi, 2014), but also against non-Muslims (Bartlett and Miller, 2012), in both cases to justify their killing. The occurrence of the keyword *kafir* within online social media posts was recently shown to be a strong indicator for jihadist hate speech (De Smedt, De Pauw and Van Ostaeysen, 2018). We use male, female and plural forms (*kafir*—*kafirah*—*kuffar*) in Arabic script for the query.

We used the ForSight platform provided by Crimson Hexagon\(^3\) to retrieve daily time series of the global Twitter volume for our keywords. We excluded posts with the keyword *RT* to ignore retweets. The time series are based on the full Twitter stream, which makes the numbers exact. Unless otherwise noted, we preprocessed the original time series by taking the logarithm to base 2 and subtracting the running mean over the past 30 days to make

\(^{3}\text{https://www.crimsonhexagon.com/}\)
them stationary.

APPENDIX B: SIMULATION STUDY

We support our theoretical results with Monte Carlo simulations that demonstrate their validity and performance.

**B.1. Comparison of the null distributions.** The central result from Section 4.2 is that under the null hypothesis of independence (and some constraints on the time series), the number of trigger coincidences for a fixed threshold approximately follows the binomial distribution from Equation 17, where the success probability is obtained from a GEV distribution. This approximate result is useful specifically for the case of time series with serial dependencies, where the Bernoulli-based null distribution from Equation 7 cannot be applied. We now demonstrate that the Bernoulli-based null distribution indeed fails to describe the empirically observed numbers of trigger coincidences for time series with serial dependencies, while our GEV-based null distribution accurately describes the observed data. For this purpose, we simulate three time series with MA orders of 0, 32 and 64. For every time series, we simulate 1,000 independent pairs of event series with $N = 32$ events, and record the numbers of trigger coincidences at the three thresholds $\tau \in \{3, 4, 5\}$, with time tolerance $\Delta = 7$. For every time series and choice of threshold, we compare the empirically obtained (Monte Carlo) null distribution with the two analytical null distributions. All cumulative probability mass functions are visualized in Figure 7. The visualizations clearly show that our GEV-based estimate closely follows the empirical distribution in all runs, while the Bernoulli-based estimate is only correct for iid time series. These results also demonstrate that a value of $\Delta = 7$ is already large enough for the GEV approximation to be valid.

**B.2. Analysis of the surprise measure.** Our second central result is the Markov chain model for multiple thresholds in Section 4.3, with the associated surprise measure that complements visual analyses of trigger coincidence processes. If an observed trigger coincidence process is significantly more surprising than typical trigger coincidence processes under the null hypothesis of independence, we reject the null hypothesis in favor of peak dependence at various magnitudes. Intuitively, a trigger coincidence process is surprising if a high fraction of events triggers exceedances of very high thresholds, i.e., if the curve is close to the top-right corner in a QTR plot. To confirm that our surprise measure captures our intuitive notion of “surprise,” we illustrate its scores on a simulated time series with MA order...
8. We choose a time tolerance of $\Delta = 7$ and fit the GEV distribution $G$ to the resulting block maxima from the time series by maximum likelihood estimation. We select 32 thresholds at equally spaced $p$-quantiles between .75 and 1 from the time series and use $G$ to obtain the parameters of the Markov chain model described by Equation 17 for the smallest threshold and Equation 23 for all larger thresholds. We then simulate 1,000 independent event series with $N_E = 32$, compute the trigger coincidence processes for all event series, and obtain their surprise values from the Markov chain model. All resulting trigger coincidence processes are plotted in Figure 8, colorized by their surprise values. We also plot the most and least surprising trigger coincidence process that are theoretically possible; we obtain them by maximizing (minimizing) the surprise measure over all possible processes with a dynamic programming approach. At last, we show the marginally expected trigger coincidence process at every threshold $\tau_m$, i.e., the value of $E[K_{\Delta,\tau_m} | N_E = 32]$ obtained from Equation 17. All simulated trigger coincidence processes are close to the marginally expected sequence; the more they tend towards the top-right corner of the plot, the more surprising they are. The theoretically most surprising trigger coincidence process closely traces the top-right corner, which corresponds to our intuitive notion of the
most unusual outcome: all events trigger exceedances of the highest quantiles.

REFERENCES


Language Processing for Social Media (SocialNLP).


---

**Data Science and Data Engineering**

b-it, University of Bonn
Endenicher Allee 19C
53115 Bonn, Germany
E-mail: scharwaechter@bit.uni-bonn.de

**Data Science and Data Engineering**

b-it, University of Bonn
Endenicher Allee 19C
53115 Bonn, Germany
E-mail: mueller@bit.uni-bonn.de