Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

A GENERATIVE STOCHASTIC GRAPHICAL MODEL FOR SIMULATING SOCIAL PROTEST

Dharmashankar Subramanian

IBM Research
1101, Kitchawan Road, Rte 134
Yorktown Heights, NY 10598, USA

Lucia L. Titus

Laboratory for Analytic Sciences, NCSU
1021 Main Campus Drive
Raleigh, NC 27606, USA

ABSTRACT

Civilian protest is a complex phenomenon where large numbers of protestors participate in demonstrations. It involves multiple groups, various trigger events and social reinforcement where groups excite each other. We present a graphical generative model in which a baseline spontaneous process may undergo excitation due to external triggers, as well as inter-group contagion. We define a trigger-conditional multivariate Hawkes process, where excitation is conditional on the presence of active triggers. An arrival in this process corresponds to a batch of protestors, and random marks on the arrival serve to capture both the excitation-related parameters as well as the size of protest. The batch arrival intensity and the batch size, while mutually independent, exhibit respective history-dependence due to memory that is modeled in the excitation phenomena. We present a simulation algorithm for generating sample paths, and results estimating likelihood of large-scale protest on a realistic model.

1 INTRODUCTION

Civilian protest is a complex phenomenon that culminates in the form of demonstrations, each with a large number of protestors. Recent civil unrest events in Brazil provide vivid examples of mass protest involving multiple groups of participants, as well as diverse triggers like transportation fare hike, police brutality, excessive government spending on stadiums, and corruption and bribery involving public officials; see (2015-16 protests in Brazil 2015). A notable aspect of mass protest is that both external factors as well as contagion-like endogenous factors play a role in its dynamics. A study of the Brazil protests (Winters and Weitz-Shapiro 2014) describes how it started as a response to bus fare hikes and organically developed into a significantly wider participation from multiple groups over a bigger set of issues such as government inefficiency and degrading social infrastructure in health and education among others.

With respect to generative models for unrest, there is existing literature that is based on viral propagation over a network. Unrest rising from social diffusion across connected spatial regions is presented in (Braha 2012). There are similar viral models with focus on Twitter data as a measurable proxy, including applications of epidemic models and activity cascades; see Hodas and Lerman (2014), Gleeson et al. (2015), Goode et al. (2015). However, these models are not applicable for quantitatively analyzing the risk of social protest over some long-term future horizon. For instance, they don't account for trigger events in the environment that may heighten the prospect of protest (Snow and Moss 2014). They ignore the possibility of multiple distinct groups of participants, and also vary in their extent of including social reinforcement within or across groups. Lastly, with respect to forecasting models of social unrest over the shorter-term, there is recent work on a fully automated system called EMBERS (Ramakrishnan et al. 2014), which senses signals from open source indicators like tweets, news, and blogs to issue predictions over a week to ten days of lead-time.

Our research in this paper is motivated by a risk analysis perspective on social protest. Our application involves a Government societal risk analyst who might analyze a longer-term future spanning several months to a year using qualitative techniques. We are interested in how to quantitatively complement such an effort. We seek to do so by developing a generative formulation that allows the analyst to use background knowledge for building models and conduct computational studies via simulation. The simulation is driven by a combination of subjective beliefs around some forward-looking model parameters, and historical data driven estimation of the remaining model parameters. The historical data that is available via curation and annotation (Titus 2016) is as follows. Over a duration, say Jan 1 2015 to Dec. 31 2016 for e.g., the data contains specific days that experienced mass protest reported in the media, along with the names of groups that participated and an estimate of the overall size of protest. The time line spanned by the data set is also annotated by analysts with the historical onset time for various triggers, and an estimate of the duration of active persistence of each such trigger. Motivated by this, we introduce a novel stochastic graphical model for social protest by building on the idea of a multivariate Hawkes process.

2 TOWARDS A GENERATIVE RISK SIMULATION MODEL OF SOCIAL UNREST

We consider the problem of modeling social protest over a chosen interval $[0,H]$ into the future and involving a set G of multiple social groups. A model for protest should ideally incorporate multiple exogenous trigger events, multiple protest groups with their sensitivities to various triggers, and contagion-like dynamics within and across groups. The first two items are typically reflected in qualitative methods such as mind maps, alternative futures analysis and scenario development (Prunckun 2010, Schwartz 1996, Center for the Study of Intelligence 2009). We present a quantitative model in this section that addresses each of these.

2.1 Exogenous Events

We consider two types of trigger events. Single-occurrence events (set S) that may happen at most once, and multi-occurrence events (set M) that may happen more than once over the horizon. The choice of the event type depends on the nature of the event in terms of its typical time-scale, and how it relates to the length of the horizon under consideration. For e.g., over a 1 year horizon, an event like a political corruption scandal in a country where public corruption is high may be treated as a multi-occurrence event, whereas an event like an economic recession may be treated as a single-occurrence event.

For each single-occurrence event, $s \in S$, we associate an occurrence probability p_s. We also include a random timing t_s that denotes the onset conditional on occurrence, and a duration d_s that denotes an interval over which the event persists. The event persistence may be thought of as a type of social memory, to model the recency effect of the trigger event with respect to influencing protest. For each multi-occurrence event, $m \in M$, we associate a Poisson model in terms of its intensity, λ_m over $[0,H]$, along with a duration d_m for each occurrence of this event. Let $\mathtt{Dist}(t_s)$, $\mathtt{Dist}(d_s)$ and $\mathtt{Dist}(d_m)$ denote distribution functions for these random variables specified by the risk analyst, as per subjective beliefs. This is reasonable (Aven and Cox 2016) because the events in the longer-term future are typically not always exchangeable with past occurrences, but the analyst's belief may however be informed by historical base rate of occurrence. For notational purposes, let $R = S \cup M$ denote the set of all exogenous risk events of interest.

2.2 Protest Groups

Firstly, we associate a Poisson process with a baseline intensity λ_g with each protest group $g \in G$, where an arrival corresponds to a protest event involving a batch of protestors. This is meant to capture the baseline, i.e. random participation in protest without any specific trigger event; for e.g. rallying in support of long-standing issues that define the group's agenda. From a generative point of view, it provides a necessary seeding process that can undergo further excitation under suitable conditions, as we describe later in the paper. Let $\mathscr{B}_g$ denote the set of baseline arrivals in any group g, where each $i \in \mathscr{B}_g$ carries a random mark μ_i that denotes the size of the event. This random mark is taken to be independent across baseline arrivals in any group, as well as across groups.

2.3 Hawkes Process And Excitation

We begin with a brief introduction to a Hawkes process (Hawkes 1971), which provides a principled way to model situations where the occurrence of a past event gives a temporary boost to the probability of an event in the future. It is a point process whose intensity is time-varying and random; it is the sum of a baseline background rate and additive contributions from past events. In a univariate setting, past events of the only arrival process under question contribute to what is known as self-excitation. In a multivariate setting, one considers a vector of such Poisson processes that may also exhibit mutual-excitation (Embrechts, Liniger, and Lin 2011), where additive contributions to any component process, say g, may result from past arrivals in any other component processes, say $k \neq g$. It has seen applications in seismology (Ogata 2013), finance (Filimonov and Sornette 2012) and tweet cascades in social networks (Cadena et al. 2015). In construction, it is like a branching process, where an immigrant event is an arrival from a baseline Poisson process. An offspring event is a realization from an offspring Poisson process that is born at the time of arrival of its parent event, per excitation. This offspring process starts with an intensity equal to an additive bump that is imparted by its parent event, and its intensity decays at some rate to zero. Further, offspring events may similarly reproduce due to excitation. Also, each offspring Poisson process is independent of its parent process, given the specific parent event that produces it and endows it with a corresponding starting intensity. See Simma and Jordan (2012) for a random forest view of a univariate Hawkes process. Notation-wise, each baseline arrival in the above view, $i \in \mathscr{B}_g$, seeds its own cluster, which we denote as $\mathscr{C}_i$. Let $\mathscr{O}_g$ denote the set of all offspring arrivals corresponding to group g. By the constructive view, each $j \in \mathscr{O}_g$ has a unique baseline arrival ancestor, $\texttt{Ancestor}(j) \in \mathscr{B}_{g'}$ from some group $g' \in G$ (potentially, $g' \neq g$). Likewise, it also has an immediate parent, $\texttt{Parent}(j) \in \mathscr{C}_{\texttt{Ancestor}(j)}$, from some group, i.e. in general a cluster has arrivals from multiple groups. It also has a size μ_j to capture the magnitude. Let I denote the set of all arrivals (baseline and offspring) across all groups, and for any $i \in I$, let $\texttt{Group}(i) \in G$ denote the group-type and T_i denote the the time of arrival.

2.4 A Graphical Model For Social Protest

Consider a graph $\mathscr{G}(\mathscr{V},\mathscr{E})$ whose node set is taken as $\mathscr{V} = R \cup G$, i.e. the set of exogenous events and protest groups. Further, let $\mathscr{E}$ consist of directed edges of the form (u,v), where $u \in R$ and $v \in G$, i.e. edges denoting the influence exerted by trigger events on groups. Additionally, let $\mathscr{E}$ also consist of directed edges of the form (u,v), where $u,v \in G$, i.e. edges denoting directed influence among protest groups. These edges represent the potential for self-excitation (self-loops) and mutual-excitation (edges with distinct vertices) across the set of Poisson processes. Let R_g denote the set of triggers that influence group g, i.e. $R_g = \{r | r \in \texttt{Predecessor}(g) \cap R\}$. Let P_g denote the set of groups whose events potentially trigger protest in group g via excitation, i.e. $P_g = \{k | k \in \texttt{Predecessor}(g) \cap G\}$. Figure 1 shows an example that has three protest groups, g_1, g_2, g_3, two single-occurrence risk events, s_1, s_2, and two multi-occurrence risk events, m_1, m_2, respectively. Past events of each group in this example have the potential to self-excite future events in their own group, whereas past events in g_1 have the potential to also excite future events in both g_2 and g_3 respectively. We begin with the following definition of a *trigger-conditional* Hawkes

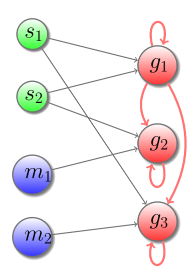

Figure 1: A graphical model with 3 protest groups.

process.

Definition 1 (Trigger-conditional Hawkes Process) A trigger-conditional multivariate Hawkes process is conventional multivariate Hawkes process with the following additional conditions. A Parent event i arriving at T_i in any component process g is eligible to reproduce, i.e. give birth to an offspring Poisson process in each downstream component $k \in \texttt{Successor}(g)$ only if there is a corresponding trigger (risk) event $r \in R_g$ that is active at T_i. Further, if multiple risk event types $r \in R_g$ are active at T_i, each unique type of active trigger enables parent i to give birth to a separate, independent offspring Poisson process in each downstream component $k \in \texttt{Successor}(g)$.

2.4.1 Marked Arrivals In Each Group

In order to model random excitation as per Definition 1, let us consider an arrival i at time T_i in any group g. Corresponding to any group $k \in \texttt{Successor}(g)$, the above arrival i carries three additional marks, $(\Delta\lambda_{g,k,r,i}, \Delta\tau_{g,k,r,i}, \Delta\mu_{g,k,r,i})$. $\Delta\lambda_{g,k,i}$ denotes the bump in the intensity of arrivals in group k due to arrival i, if so enabled as per Definition 1. $\Delta\tau_{g,k,r,i}$ denotes the duration of time over which such a bump in intensity will decay to zero. So a child Poisson process in group k with intensity $\Delta\lambda_{g,k,i}$ is conditionally born at time T_i, and this child process will decay to zero intensity over $[T_i, T_i + \Delta\tau_{g,k,r,i}]$. Lastly, $\Delta\mu_{g,k,r,i}$ denotes amplification in the size of excited protest, relative to baseline protest size. All random marks are assumed to be independent across arrivals, as well as mutually independent within any given arrival. Let $(\texttt{Dist}(\Delta\lambda_{g,k,r}), \texttt{Dist}(\Delta\tau_{g,k,r}))$ denote distribution functions associated with each edge (g,k) with respect to $r \in R_g$. Further, we specifically restrict the model to a linear decay function Φ for bump in intensity from any specific past event, as given below, where $\Delta\lambda$ and $\Delta\tau$ denote a representative bump in intensity and duration of decay respectively, and $[x]^+ = max(x,0)$. Φ is zero by definition for $t < 0$, and is given as

$$\Phi(t, \Delta\lambda, \Delta\tau) = \left[\Delta\lambda - \frac{t\Delta\lambda}{\Delta\tau}\right]^+ \quad (\text{for } t \geq 0). \tag{1}$$

2.4.2 Net Arrival Rate In Each Group

With the above setup, we may express the net time-varying rate, $\nu_g(t)$, in any group g as per Definition 1. Let $I_p = \mathscr{B}_p \cup \mathscr{O}_p$ denote the set of all arrivals in group p. We have that,

$$\nu_g(t) = \lambda_g + \sum_{p \in P_g} \sum_{r \in R_p} \sum_{i \in I_p | T_i < t} \mathbb{1}_{r,T_i} \Phi(t - T_i, \Delta\lambda_{p,g,r,i}, \Delta\tau_{p,g,r,i}). \tag{2}$$

It may be seen via Equation 2 that the trigger-conditional Hawkes process behaves like a hybrid process in each component g which may switch between a homogeneous Poisson process with a baseline intensity and a history-dependent Poisson process with time-varying intensity. The indicator function $\mathbb{1}_{r,T_i}$ is a binary variable to indicate whether the risk event r is active in its persistence at time T_i. If $r \in S$, i.e. a single-occurrence risk event (as per notation defined in Section 2.1),

$$\mathbb{1}_{r,T_i} = \begin{cases} 1 & \text{if } t_r \leq T_i \leq t_r + d_r, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

If $r \in M$, i.e. a multi-occurrence risk event, let T_k^r denote the onset times of multiple occurrences (indexed by k). The indicator function $\mathbb{1}_{r,T_i}$ in this case is a binary variable that indicates whether there is at least one such arrival of the risk event that is active in its persistence at time T_i. In this case (as per notation defined in Section 2.1),

$$\mathbb{1}_{r,T_i} = \max_k (\mathbb{1}_{T_k^r, T_i}), \tag{4}$$

where we have that

$$\mathbb{1}_{T_k^r, T_i} = \begin{cases} 1 & \text{if } T_k^r \leq T_i \leq T_k^r + d_k, \\ 0 & \text{otherwise.} \end{cases}$$

Lastly, we model a multiplicative amplification over and above a baseline magnitude for the size of all excited arrivals as,

$$\mu_j = \tilde{\mu}(1+\Delta\mu_{k,g,r,j}) \quad \text{(where } k = \texttt{Group}(\texttt{Parent}(j))\text{)}, \tag{5}$$

where the amplification is with respect to a latent random variable that is distributed as $\tilde{\mu} \sim \texttt{Dist}(\mu_g)$.

3 SIMULATING THE GENERATIVE MODEL

We address the simulation of the process described by Equation 2 to generate sample paths of protest. We consider a simulation horizon $[0,H]$ consisting of a discrete set $\{1,\ldots,H\}$ of H time periods in some appropriate unit of time (say, weeks). Each continuous-time Poisson arrival over $[0,H]$ would then occur within one of the discrete time periods. The main idea is to recognize that Equation 2 presents a additive decomposition of the net arrival rate in terms of contributions that conditionally appear on the right hand side. If the net Poisson arrival intensity is due to an additive set of contributions from mutually independent contributing Poisson processes, then we may simulate each independent contribution individually on a common time line and superpose the arrivals by placing them together on the same common time line. This is the superposition property of Poisson processes in the mutually independent setting (Ross 2014). While Equation 2 appears more complicated with historical dependence on a random number of contributing processes, these contributions are mutually independent given their respective parent events.

We adapt the above essential idea by taking a random forest view of the process. So instead of directly working with Equation 2, we instead focus the simulation on each cluster $\mathscr{C}_i$ that is seeded by some baseline arrival $i \in \mathscr{B}_g$ in some group g, as defined in Section 2.3. This is related to the approximate simulation approach presented for a univariate Hawkes process in Møller and Rasmussen (2006). Recall that each cluster is a rooted tree that captures a set of arrivals, possibly spanning multiple groups, with a parent-child relationship across successive levels in the tree. Starting at $t = 0$, we firstly generate a set of immigrant arrivals for each group using the corresponding baseline Poisson processes. We maintain an active list of unprocessed immigrant arrivals in an ascending order of their arrival times. We sequentially process the active list to simulate a cluster corresponding to each element in the list, noting that each immigrant is the root of a cluster that it spawns. Within each cluster, we recursively simulate arrivals from offspring Poisson processes that descend from previous simulated arrivals in the cluster, and which are conditionally enabled as per Equations 3 and 4. We do so by maintaining an active list of unprocessed offspring Poisson processes in the current cluster, organized in an ascending order of their birth times. For e.g., starting with a baseline immigrant, say i in group g, we simulate arrivals from offspring Poisson processes in groups $k \in \texttt{Successor}(g)$ if they are enabled as per Equations 3 and 4, and continue so recursively with further generations of offspring Poisson processes if enabled, and so on. If we think of Equation 2 as a set of equations indexed by group, the cluster-centric simulation captures the cluster-relevant contributions from the full set of right hand sides in Equation 2. Superposing the simulated set of arrivals across all simulated clusters on a common time line essentially produces a simulation corresponding to the set of arrival processes in Equation 2. Lastly, we assume that the baseline Poisson intensity in each group is $\lambda_g \mathbb{1}_{t\geq 0}$, i.e. we ignore edge-effect possibility of immigrant arrivals with arrival times that are before $t = 0$, the simulation start time. This essentially ignores the possibility of offspring arrivals in $[0,H]$ stemming from such historical parent events.

Algorithm 1 presents the full procedure. The sampling steps from various probability distribution functions have standard techniques (see for e.g. Ross (2012), Law and Kelton (2000)) and implementations (see for e.g. Jones, Oliphant, Peterson, et al. (2001)). Techniques to sample arrivals from homogenous Poisson processes may be found in Pasupathy (2011a), while those for nonhomogenous Poisson processes, such as our time-varying Poisson intensity Φ from Equation 1, may be seen in Pasupathy (2011b). It should be noted that the branching process view of the generative model also allows the possibility of a run-away in the simulation, i.e. a situation where the number of offspring grows indefinitely and the inner `while` loop in Step 3 continues without termination. This may happen when the various excitation parameters in the model, namely, $\Delta\lambda$ and $\Delta\tau$ assume sufficiently large values and the average number of offspring stemming from a parent event is greater than one. Such a regime is referred to as super-critical or explosive (Filimonov and Sornette 2015) from the view of a branching process. In order to address this in practice, we suggest that the termination condition of the inner `while` loop consider alternative criteria. For e.g. one may consider the earlier of two criteria, namely that of reaching a suitably large upper limit on the number of unprocessed offspring in the corresponding ACTIVECHILDREN list, or as stated in Algorithm 1, the criterion of emptying out the corresponding ACTIVECHILDREN list. The former case represents the run-away condition, which may be interpreted by the analyst as the realization of a very large protest participation. A more nuanced simulation procedure may also choose to handle the run-away condition via an upper limit on the resulting size of participation in the corresponding group, as follows. Recall that we use a discrete-time model for keeping track of the number of arrivals and their corresponding protest sizes. So we may choose to ignore future offspring arrivals in time-periods that have already reached an upper limit on the size of protest participation from the corresponding group (due to already processed arrivals). Such an upper limit may be provided by the analyst as an estimated bound on the size of each protest group.

For each simulated time line, we may capture various quantities of interest to support risk analyst queries. These include a) the cluster-specific view in terms of the seeding lineage of historical events along with their respective group identities, arrival times and protest sizes, as well as b) the trigger events that were active at the time of each parent event along with the onset and duration of persistence of each trigger event. Mapping the continuous arrival times of various arrivals to the discrete time scale over $[0,H]$, we may generate a cumulative view of the protest size by group as well as across groups over the horizon of interest at the desired time resolution for the risk analyst (e.g. weekly). Similarly, one may also track the composition of protestors in terms of their generation index $(0^{th}, 1^{st}, \ldots)$ from any given group that protested in any given time period to gain insights about the contagion dynamics. Such fine-grained book-keeping would enable informative responses to queries from the risk analyst such as: *Tell me more about the history and dynamics that led to the large cumulative protest that I see in this simulated time line at time period, say T.* The collective implications of the analyst's subjective beliefs about the future, as well as historical data, are exposed via such queries whose responses may serve insights around risk preparedness with respect to risk consequences.

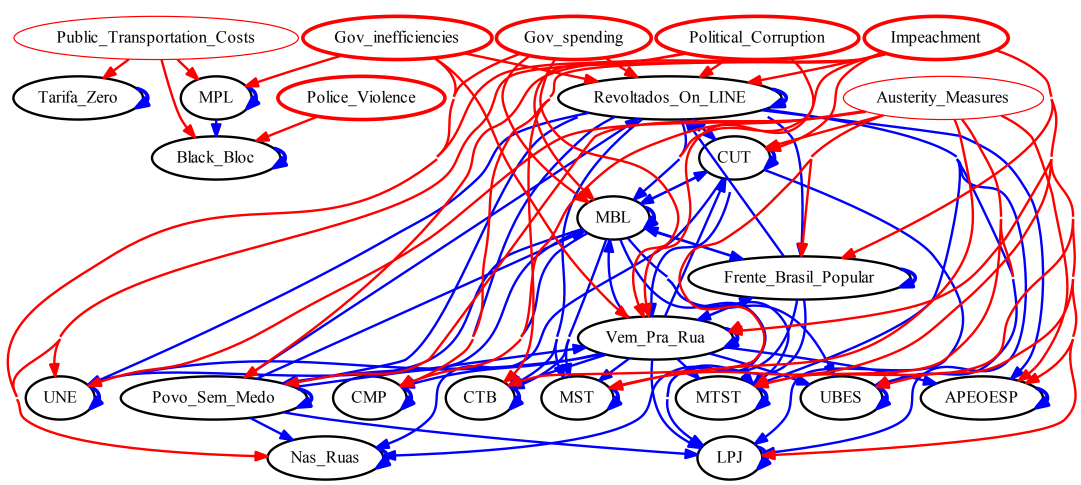

Figure 2: A graphical model with 18 protest groups

4 COMPUTATIONAL RESULTS

We present computational results on a model instance based on protest groups in Brazil, as shown in Figure 2. The model involves 18 protest groups shown in black (outline) nodes, 5 multi-occurrence risk

Algorithm 1 Trigger-conditional Social Protest Generative Process simulation

specify
  Number of simulations NUMSIM
  Horizon H, Graphical model $\mathscr{G}(\mathscr{V},\mathscr{E})$
  $\texttt{Dist}(t_s)$, $\texttt{Dist}(d_s)$, p_s for each $s \in S$ and $\texttt{Dist}(d_m)$, λ_m for each $m \in M$
  λ_g, $\texttt{Dist}(\mu_g)$ for each $g \in G$
  $\texttt{Dist}(\Delta\lambda_{g,k,r})$,$\texttt{Dist}(\Delta\tau_{g,k,r})$,$\texttt{Dist}(\Delta\mu_{g,k,r})$, $\forall(g,k) \in \mathscr{E}$, where $g \in G$, $k \in \texttt{Successor}(g)$, and $r \in R_g$
repeat
  0. **Initialize**
    COUNTER $= 0$,
    ACTIVEIMMIGRANTS $\leftarrow$ EMPTY SORTEDARRIVALLIST (list sorted by arrival time)
    $\mathscr{B}_g = \varnothing$, $\mathscr{O}_g = \varnothing$, $\forall g \in G$
  1. Generate risk event realizations for each $r \in R$
    For each single-occurrence event $s \in S$, sample $\delta_s \sim$ Bernoulli(p_s), and associate:
      $t_s \sim \texttt{Dist}(t_s)$, $d_s \sim \texttt{Dist}(d_s)$
    For each multi-occurrence event $m \in M$, for arrival $k \sim$ Poisson(λ_m) at time T_k^m associate:
      $t_{i_m} \sim \texttt{Dist}(t_m)$, $d_{i_m} \sim \texttt{Dist}(d_m)$
  2. Generate a set of immigrant arrivals i_g $\forall g \in G$ and store them in ACTIVEIMMIGRANTS
    GENARRIVALS(Poisson(λ_g), ACTIVEIMMIGRANTS, NULL, g, 0)
  3. Cluster Simulation
    ▷ Simulate a cluster seeded by each immigrant from above ACTIVEIMMIGRANTS
      while ACTIVEIMMIGRANTS is not empty **do**
      ▷ POP below removes the item at the head of the sorted list, i.e. least arrival time item
        $i =$ ACTIVEIMMIGRANTS $\leftarrow$ POP, $g = \texttt{Group}(i)$, $\mathscr{B}_g = \mathscr{B}_g \cup i$
        ACTIVECHILDREN(i) $\leftarrow$ EMPTY SORTEDARRIVALLIST
        for each $r \in R_g$ **do**
          if $(\mathbb{1}_{r,T_i} = 1)$ (as per Equations 3, 4)
            for each $k \in \texttt{Successor}(g)$ **do**
            ▷ Generate offspring i_k and PUSH to ACTIVECHILDREN(i)
              GENARRIVALS(Poisson$(\Phi(T_i,\Delta\lambda_{g,k,r,i},\Delta\tau_{g,k,r,i}))$,ACTIVECHILDREN$(i)$, i, k, T_i)
              (see Equation 1 for $\Phi(t,\Delta\lambda,\Delta\tau)$))
            end for
        end for
        while ACTIVECHILDREN(i) is not empty **do**
          $j =$ ACTIVECHILDREN(i) $\leftarrow$ POP, and $f = \texttt{Group}(j)$, $\mathscr{O}_f = \mathscr{O}_f \cup j$
          for each $r \in R_f$ **do**
            if $(\mathbb{1}_{r,T_j} = 1)$
              for each $k \in \texttt{Successor}(f)$ **do**
              ▷ Generate offspring i_k and PUSH to ACTIVECHILDREN(i)
                GENARRIVALS(Poisson$(\Phi(T_j,\Delta\lambda_{f,k,r,j},\Delta\tau_{f,k,r,j}))$,ACTIVECHILDREN$(i)$, j, k, T_j)
              end for
          end for
        end while
      end while
  ▷ Unioning all arrivals in the resulting $\mathscr{B}_g$ and $\mathscr{O}_g$ across all groups $g \in G$, and superposing them on a common timeline, contains the simulated trace of all arrivals in the process described by Equation 2.
  4. COUNTER $=$ COUNTER $+ 1$
until COUNTER $<$ NUMSIM

Procedure 1 GENARRIVALS

specify Horizon H

GENARRIVALS(Poisson(γ), LIST, PARENT j , GROUP g, START-TIME T_0)

Generate arrivals $i \sim$ Poisson(γ) over remaining time $[T_0, H - T_0]$ and store them in LIST

▷ For each arrival $i \sim$ Poisson(γ) arriving at time T_i, associate:

▷ Group and Cluster-related information: $\texttt{Group}(i) = g$

if j is NULL

$\texttt{Ancestor}(i) = j$, $\mu_i \sim \texttt{Dist}(\mu_g)$

else

$\texttt{Parent}(i) = j$, $\texttt{Ancestor}(i) = \texttt{Ancestor}(j)$, μ_i as per Equation 5

▷ Excitation-related information: $\forall k \in \texttt{Successor}(g)$ and $\forall r \in R_g$

$\Delta\lambda_{g,k,r,i} \sim \texttt{Dist}(\Delta\lambda_{g,k,r})$, $\Delta\tau_{g,k,r,i} \sim \texttt{Dist}(\Delta\tau_{g,k,r})$, $\Delta\mu_{g,k,r,i} \sim \texttt{Dist}(\Delta\mu_{g,k,r})$

Immigrants i sorted in an ascending order of arrival times T_i in LIST

factors shown in bolded red nodes and 2 single-occurrence risk factors shown in red nodes, and its structure was built in collaboration with government analysts. The red edges denote the triggering influences exerted by the risk factors on the protest groups, and the blue edges denote excitations in the generative model formulation defined in Section 2.4. We consider the model over a horizon $H = 16$ weeks. The excitation phenomenon is parameterized at the resolution of a hyperedge involving three nodes, namely a risk trigger node and two excitable nodes that correspond to the same (self-excitation) or two distinct (cross-excitation) protest groups. Table 1 shows all such hyperedges that exhibit non-trivial cross excitation, which serves as a transfer mechanism, i.e. triggers that don't directly affect a group may indirectly excite it via another group that they directly affect. Table 3 shows all the self-excitation combinations, where the first column stands for both the source and target groups, and the second column shows the triggers. The remaining columns show the excitation related parameters that are assumed to hold for both self- and cross-excitations. Parameters with minimum, mode and maximum are taken as random variables with a triangular density with these values, whereas those with minimum and maximum alone are taken as uniform random variables over the specified range. All parameters are estimated from historical data Titus (2016), and we omit the details due to space, and details may be found in Subramanian (2016). The parameters corresponding to the futuristic beliefs of the analyst around risk triggers is assumed to be as per Table 4 and Table 2.

Figure 3 shows five sample paths in solid lines (out of 10000 projections) over 16 weeks that are generated via simulation by the model. It also shows three dashed lines corresponding to three different (and non-overlapping) intervals of 16 week durations taken from the historical data set. As seen in the figure, the generative model is rich enough to capture the range of actual protest trajectories that were witnessed in the data set. Paths 1 and 2 exceed 10 Million at their peak in aggregate number of protestors, whereas Path 5 witnesses a maximum of 20000 and appears close to zero due to the scale of the y-axis in the figure. With more simulations, one may explore further extreme scenarios (such as the dashed line 'Actual 1') that are potentially within the reach of the model. The figure shows a simulation based estimate of the likelihood of exceeding 1 Million of aggregate number of protestors in at least one week over the horizon. The annotations on the plot show the identities of the protest groups that were dominant participants in these sample paths, along with the triggers that were active over respective intervals of time (shown as a weekly range in square brackets). For e.g. in Path 2, 'Political Corruption' trigger was active over weeks 7-10, 'Public Transportation' trigger was active over weeks 8-11, and the dominant participation was from Vem Pra Rua, Frente Brasil Popular and Revoltados.

5 CONCLUSIONS

We have presented a graphical, generative risk model for simulating social protest involving a network of multiple groups and external triggers. Our model integrates three distinct phenomena to generate protest,

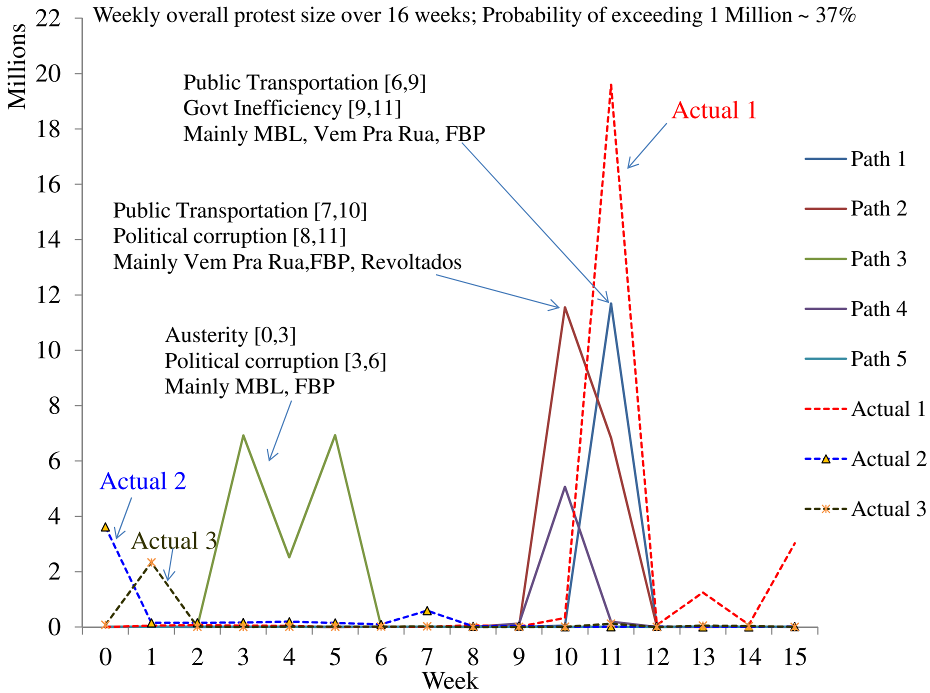

Figure 3: Simulated projections of aggregate weekly protest

namely, a baseline spontaneous phenomenon that is due to apparent randomness, an exogenously excited phenomenon that is due to external events in the environment, and an endogenously excited phenomenon that is due to a stochastic contagion effect both within and across protest groups. We have developed a novel formulation by defining a *trigger-conditional* adaptation of multivariate Hawkes processes, which exhibit self-excitation and mutual excitation that is further conditional on the presence of active triggers. An arrival in this process corresponds to a batch of protestors, and random *marks* on the arrival serve to capture both the excitation-related parameters as well as the size of protest. Both the batch arrival intensity and the batch size, while mutually independent, exhibit respective history-dependence due to memory that is modeled in the excitation phenomena. We also present a simulation algorithm for generating sample paths.

In a productive collaboration with the analysts, we identified and assembled historical data with necessary annotations in line with the needs of a realistic model instance. We performed data analysis for estimating various model parameters and have reported computational results from our model with these estimates. The generative capacity of the proposed stochastic model is sufficiently rich to capture a realistic range of protest event trajectories and participation volumes. The model itself is more general than its inspiration, namely the phenomenon of social protest, and may be used in any context where one seeks to model contagion that is due to external triggers and network interaction effects. The proposed *trigger-conditional* multivariate Hawkes process provides a natural format to mathematically express such dynamics.

ACKNOWLEDGMENTS

This material is based upon work supported in whole or in part with funding from the Laboratory for Analytic Sciences (LAS). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the LAS and/or any agency or entity of the United States Government.

Table 1: Edge-risk combinations that are non-trivial in the model.

Source	Target	Trigger	Source	Target	Trigger
Revoltados On LINE	APEOESP	Gov inefficiencies	Povo Sem Medo	MBL	Austerity Measures
Revoltados On LINE	APEOESP	Gov spending	Revoltados On LINE	MBL	Gov inefficiencies
Revoltados On LINE	APEOESP	Political Corruption	Vem Pra Rua	MBL	Gov inefficiencies
Vem Pra Rua	APEOESP	Gov inefficiencies	MBL	MST	Gov spending
Vem Pra Rua	APEOESP	Gov spending	MBL	MST	Political Corruption
Vem Pra Rua	APEOESP	Political Corruption	Revoltados On LINE	MST	Gov inefficiencies
MPL	Black Bloc	Gov inefficiencies	Revoltados On LINE	MST	Gov spending
MBL	CMP	Gov spending	Revoltados On LINE	MST	Political Corruption
MBL	CMP	Political Corruption	Vem Pra Rua	MST	Gov inefficiencies
Revoltados On LINE	CMP	Gov inefficiencies	Vem Pra Rua	MST	Gov spending
Revoltados On LINE	CMP	Gov spending	Vem Pra Rua	MST	Political Corruption
Revoltados On LINE	CMP	Political Corruption	MBL	MTST	Gov spending
Vem Pra Rua	CMP	Gov inefficiencies	MBL	MTST	Political Corruption
Vem Pra Rua	CMP	Gov spending	Revoltados On LINE	MTST	Gov inefficiencies
Vem Pra Rua	CMP	Political Corruption	Revoltados On LINE	MTST	Gov spending
MBL	CTB	Gov spending	Revoltados On LINE	MTST	Political Corruption
MBL	CTB	Political Corruption	Vem Pra Rua	MTST	Gov inefficiencies
Revoltados On LINE	CTB	Gov inefficiencies	Vem Pra Rua	MTST	Gov spending
Revoltados On LINE	CTB	Gov spending	Vem Pra Rua	MTST	Political Corruption
Revoltados On LINE	CTB	Political Corruption	CUT	Nas Ruas	Austerity Measures
Vem Pra Rua	CTB	Gov inefficiencies	Frente Brasil Popular	Nas Ruas	Austerity Measures
Vem Pra Rua	CTB	Gov spending	Povo Sem Medo	Nas Ruas	Austerity Measures
Vem Pra Rua	CTB	Political Corruption	MBL	Povo Sem Medo	Gov spending
MBL	CUT	Gov spending	MBL	Povo Sem Medo	Political Corruption
MBL	CUT	Political Corruption	Revoltados On LINE	Povo Sem Medo	Gov inefficiencies
Revoltados On LINE	CUT	Gov inefficiencies	Revoltados On LINE	Povo Sem Medo	Gov spending
Revoltados On LINE	CUT	Gov spending	Revoltados On LINE	Povo Sem Medo	Political Corruption
Revoltados On LINE	CUT	Political Corruption	Vem Pra Rua	Povo Sem Medo	Gov inefficiencies
Vem Pra Rua	CUT	Gov inefficiencies	Vem Pra Rua	Povo Sem Medo	Gov spending
Vem Pra Rua	CUT	Gov spending	Vem Pra Rua	Povo Sem Medo	Political Corruption
Vem Pra Rua	CUT	Political Corruption	CUT	Revoltados On LINE	Austerity Measures
MBL	Frente Brasil Popular	Gov spending	Frente Brasil Popular	Revoltados On LINE	Austerity Measures
MBL	Frente Brasil Popular	Political Corruption	Povo Sem Medo	Revoltados On LINE	Austerity Measures
Revoltados On LINE	Frente Brasil Popular	Gov inefficiencies	MBL	UBES	Gov spending
Revoltados On LINE	Frente Brasil Popular	Gov spending	MBL	UBES	Political Corruption
Revoltados On LINE	Frente Brasil Popular	Political Corruption	Revoltados On LINE	UBES	Gov inefficiencies
Vem Pra Rua	Frente Brasil Popular	Gov inefficiencies	Revoltados On LINE	UBES	Gov spending
Vem Pra Rua	Frente Brasil Popular	Gov spending	Revoltados On LINE	UBES	Political Corruption
Vem Pra Rua	Frente Brasil Popular	Political Corruption	Vem Pra Rua	UBES	Gov inefficiencies
CUT	LPJ	Austerity Measures	Vem Pra Rua	UBES	Gov spending
Frente Brasil Popular	LPJ	Austerity Measures	Vem Pra Rua	UBES	Political Corruption
MBL	LPJ	Gov spending	MBL	UNE	Gov spending
MBL	LPJ	Political Corruption	MBL	UNE	Political Corruption
Povo Sem Medo	LPJ	Austerity Measures	Revoltados On LINE	UNE	Gov inefficiencies
Revoltados On LINE	LPJ	Gov inefficiencies	Revoltados On LINE	UNE	Gov spending
Revoltados On LINE	LPJ	Gov spending	Revoltados On LINE	UNE	Political Corruption
Revoltados On LINE	LPJ	Political Corruption	Vem Pra Rua	UNE	Gov inefficiencies
Vem Pra Rua	LPJ	Gov inefficiencies	Vem Pra Rua	UNE	Gov spending
Vem Pra Rua	LPJ	Gov spending	Vem Pra Rua	UNE	Political Corruption
Vem Pra Rua	LPJ	Political Corruption	CUT	Vem Pra Rua	Austerity Measures
CUT	MBL	Austerity Measures	Frente Brasil Popular	Vem Pra Rua	Austerity Measures
Frente Brasil Popular	MBL	Austerity Measures	Povo Sem Medo	Vem Pra Rua	Austerity Measures

Table 2: Single-occurrence risk triggers in the model.

Trigger, $s \in S$	Probability, p_s	Timing, $t_{s,min}$, week	Timing, $t_{s,mode}$, week	Timing, $t_{s,max}$, week	$d_{s,min}$, weeks	$d_{s,max}$, weeks
Austerity Measures	0.01	0	None	4	2	4
Public Transportation Costs	0.01	0	6	16	2	4

Table 3: Edge-risk combinations for self-excitation and excitation parameters.

Group	Triggers	$\Delta\lambda_{\cdot,g,\cdot}$ (min,mode,max)%	$\Delta\mu_{\cdot,g,\cdot}$ (min,mode,max)%	$\Delta\tau_{\cdot,g,\cdot}$ (min,max) weeks
APEOESP	Aust. Meas.,Impeach.	(0,94,1031)	(156, 3760, 5724)	(1 , 2)
Black Bloc	Polic. Viol.,Pub. Transp. Costs	(0,720,1950)	(500, 632, 640)	(1 , 2)
CMP	Aust. Meas.,Impeach.	(0,217,2423)	(297, 413, 588)	(1 , 2)
CTB	Aust. Meas.,Impeach.	(0,155,681)	(0, 792, 1030)	(1 , 2)
CUT	Aust. Meas.,Impeach.	(0,99,193)	(85, 716, 1324)	(1 , 2)
Frente Brasil Popular	Aust. Meas.,Impeach.	(0,722,741)	(67, 6568, 80191)	(1 , 2)
LPJ	Impeach.	(0,107,1212)	(485, 2562, 5779)	(1 , 2)
MBL	Gov. spend.,Impeach.,Polit. Corr.	(0,549,925)	(165, 22709, 67802)	(1 , 2)
MST	Aust. Meas.,Impeach.	(0,73,218)	(108, 644, 1658)	(1 , 2)
MTST	Aust. Meas.,Impeach.	(0,98,349)	(196, 2169, 5501)	(1 , 2)
MPL	Gov. ineff.,Pub. Transp. Costs	(0,99,583)	(180, 1349, 2900)	(1 , 2)
Nas Ruas	Gov. ineff.,Gov. spend.,Impeach.,Polit. Corr.	(0,132,811)	(179, 1266, 2399)	(1 , 2)
Povo Sem Medo	Aust. Meas.,Impeach.	(0,674,1212)	(82, 6689, 37078)	(1 , 2)
Revoltados On LINE	Gov. ineff.,Gov. spend.,Impeach.,Polit. Corr.	(0,707,4586)	(0, 70811, 152168)	(1 , 2)
Tarifa Zero	Pub. Transp. Costs	(0,185,583)	(0, 15, 15)	(1 , 2)
UBES	Aust. Meas.,Impeach.	(0,121,1071)	(143, 3773, 4931)	(1 , 2)
UNE	Aust. Meas.,Impeach.	(0,62,700)	(2152, 3064, 3501)	(1 , 2)
Vem Pra Rua	Gov. ineff.,Gov. spend.,Impeach.,Polit. Corr.	(0,1423,1462)	(249, 17990, 149800)	(1 , 2)

Table 4: Multi-occurrence risk triggers in the model.

Trigger,$m \in M$	Mean Rate μ_m per year	Duration $d_{m,min}$, weeks	$d_{m,max}$, weeks
Gov inefficiencies	0.5	2	4
Gov spending	0.5	2	4
Impeachment	0.5	2	4
Police Violence	0.5	1	2
Political Corruption	0.5	2	4

REFERENCES

2015-16 protests in Brazil 2015. "2015-16 protests in Brazil — Wikipedia, The Free Encyclopedia". Accessed April 05, 2016. https://en.wikipedia.org/wiki/2015-16_protests_in_Brazil.

Aven, T., and L. A. Cox. 2016. "National and Global Risk Studies: How Can the Field of Risk Analysis Contribute?". *Risk Analysis* 36 (2): 186–190.

Braha, D. 2012, 10. "Global Civil Unrest: Contagion, Self-Organization, and Prediction". *PLoS ONE* 7 (10): 1–9.

Cadena, J., G. Korkmaz, C. J. Kuhlman, A. Marathe, N. Ramakrishnan, and A. Vullikanti. 2015. "Forecasting social unrest using activity cascades". *PloS one* 10 (6): e0128879.

Center for the Study of Intelligence 2009. "A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis". *CIA Center for the Study of Intelligence*.

Embrechts, P., T. Liniger, and L. Lin. 2011, 08. "Multivariate Hawkes processes: an application to financial data". *J. Appl. Probab.* 48A:367–378.

Filimonov, V., and D. Sornette. 2012. "Quantifying reflexivity in financial markets: Toward a prediction of flash crashes". *Physical Review E* 85 (5): 056108.

Filimonov, V., and D. Sornette. 2015. "Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data". *Quantitative Finance* 15 (8): 1293–1314.

Gleeson, J. P., K. P. O'Sullivan, R. A. Baños, and Y. Moreno. 2015. "Determinants of Meme Popularity". *arXiv preprint arXiv:1501.05956*.

Goode, B. J., S. Krishnan, M. Roan, and N. Ramakrishnan. 2015, 10. "Pricing a Protest: Forecasting the Dynamics of Civil Unrest Activity in Social Media". *PLoS ONE* 10 (10): 1–25.

Hawkes, A. G. 1971. "Point Spectra of Some Mutually Exciting Point Processes". *Journal of the Royal Statistical Society. Series B (Methodological)* 33 (3): 438–443.

Hodas, N., and K. Lerman. 2014. "The Simple Rules of Social Contagion". *Scientific Reports* 4 (4343).

Jones, Eric and Oliphant, Travis and Peterson, Pearu and others 2001. "SciPy: Open source scientific tools for Python". [Online; accessed 2016-06-15].

Law, A. M., and W. D. Kelton. 2000. *Simulation Modeling & Analysis*. 3rd ed. New York: McGraw-Hill, Inc.

Møller, J., and J. G. Rasmussen. 2006. "Approximate Simulation of Hawkes Processes". *Methodology and Computing in Applied Probability* 8 (1): 53–64.

Ogata, Y. 2013. "A prospect of earthquake prediction research". *Statistical Science* 28 (4): 521–541.

Pasupathy, R. 2011a. "Generating homogenous Poisson processes". In *Wiley Encyclopedia of Operations Research and Management Science*. Wiley.

Pasupathy, R. 2011b. "Generating nonhomogenous Poisson processes". In *Wiley Encyclopedia of Operations Research and Management Science*. Wiley.

Prunckun, H. 2010. *Handbook of scientific methods of inquiry for intelligence analysis*. Lanham: Scarecrow Press.

Ramakrishnan, N., P. Butler, S. Muthiah, N. Self, R. Khandpur, P. Saraf, W. Wang, J. Cadena, A. Vullikanti, G. Korkmaz, C. Kuhlman, A. Marathe, L. Zhao, T. Hua, F. Chen, C. T. Lu, B. Huang, A. Srinivasan, K. Trinh, L. Getoor, G. Katz, A. Doyle, C. Ackermann, I. Zavorin, J. Ford, K. Summers, Y. Fayed, J. Arredondo, D. Gupta, and D. Mares. 2014. "'Beating the News' with EMBERS: Forecasting Civil Unrest Using Open Source Indicators". In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, 1799–1808. New York, NY, USA: ACM.

Ross, S. M. 2012, 11. *Simulation, Fifth Edition*. 5 ed. Academic Press.

Ross, S. M. 2014, 2. *Introduction to Probability Models, Eleventh Edition*. 11 ed. Academic Press.

Schwartz, P. 1996, 4. *The Art of the Long View: Planning for the Future in an Uncertain World*. Reprint ed. Currency Doubleday.

Simma, A., and M. I. Jordan. 2012. "Modeling events with cascades of Poisson processes". *arXiv preprint arXiv:1203.3516*.

Snow, D. A., and D. M. Moss. 2014. "Protest on the fly toward a theory of spontaneity in the Dynamics of Protest and Social Movements". *American Sociological Review*:0003122414554081.

Subramanian, D. 2016. "Appendix with details on data analysis". http://researcher.watson.ibm.com/researcher/files/us-dharmash/appendix.pdf.

Titus, L. 2016. "Social unrest movements database in Excel". Technical Report 16-1, Laboratory of Analytical Sciences, North Carolina State University, Raleigh, North Carolina.

Winters, M. S., and R. Weitz-Shapiro. 2014. "Partisan protesters and nonpartisan protests in brazil". *Journal of Politics in Latin America* 6 (1): 137–150.

AUTHOR BIOGRAPHIES

Dharmashankar Subramanian is a Principal Research Staff Member with IBM Research. His research interests include decision-making under uncertainty, risk analysis, and applications of statistical machine learning and mathematical modeling in a diverse set of domains. His email address is dharmash@us.ibm.com.

Lucia L. Titus is a researcher with the Laboratory for Analytic Sciences, where she collaborates on a variety of open source intelligence projects. Her research interests include social movements and social unrest, extremism and radicalization, and threat and risk assessments. Her email address is lltitus@ncsu.edu.