Fake News Mitigation via Point Process Based Intervention

Mehrdad Farajtabar [1] **Jiachen Yang** [1] **Xiaojing Ye** [2] **Huan Xu** [3] **Rakshit Trivedi** [1] **Elias Khalil** [1] **Shuang Li** [3] **Le Song** [1] **Hongyuan Zha** [1]

Abstract

We propose the first multistage intervention framework that tackles fake news in social networks by combining reinforcement learning with a point process network activity model. The spread of fake news and mitigation events within the network is modeled by a multivariate Hawkes process with additional exogenous control terms. By choosing a feature representation of states, defining mitigation actions and constructing reward functions to measure the effectiveness of mitigation activities, we map the problem of fake news mitigation into the reinforcement learning framework. We develop a policy iteration method unique to the multivariate networked point process, with the goal of optimizing the actions for maximal total reward under budget constraints. Our method shows promising performance in real-time intervention experiments on a Twitter network to mitigate a surrogate fake news campaign, and outperforms alternatives on synthetic datasets.

1. Introduction

The recent proliferation of malicious fake news in social media has been a source of widespread concern. Given that more than 62% of U.S. adults turn to social media for news, with 18% doing so often, fake news can have potential real-world consequences on a large scale (Gottfried & Shearer, 2016). For example, within the final three months of the 2016 U.S. presidential election, news stories that favored either of the two nominees–later proved to be fake–were shared over 37 million times on Facebook alone, and over half of those who recalled seeing fake news stories believed them (Allcott & Gentzkow, 2017). An analysis by Buzzfeed News shows that the top 20 false election stories from hoax websites generated nearly 1.5 million more user engagement activities on Facebook than the top 20 stories from reputable major news outlets (Silverman, 2016). Therefore, there is an urgent call to develop effective rectifying strategies to mitigate the impact of fake news.

Policies to counter fake news can be categorized by the level of manual oversight and the aggressiveness of action required. Aggressively acting on fake news has various drawbacks. For example, Facebook's strategy allows users to report stories as potential fake news, sends these stories to fact-checking organizations, and flags them as disputed in users' newsfeed (Mosseri, 2016). Such direct action on the offending news requires a high degree of human oversight, which can be costly and slow and also may violate civil rights. The report-and-flag mechanism is also open to abuse by adversaries who maliciously report real news. Given these disadvantages, we consider an alternative strategy: optimizing the performance of real news propagation over the network. Intuitively, we want people's exposure to real news to match their exposure to fake news.

We face several key modeling and computational issues. For example, how to quantify the uncertainty of user activities and news propagation within the network? How to measure the effect of mitigation incentives and activities? Is it possible to steer the spontaneous user mitigation activities by an intervention strategy? To address these questions, we model the temporal randomness of fake news and mitigation events ("valid news") as multivariate point processes with self and mutual excitations, in which the control incentivizes more spontaneous mitigation events by contributing to the exogenous activity of campaigner nodes. The influence of fake news and mitigation activities is quantified using event exposure counts (i.e. the number of times that a user is exposed to fake or real news posts from other users whom she follows).

Our key contributions are as follows. We present the first formulation of fake news mitigation as the problem of optimal point process intervention in a network. The goal is to optimize the activity policy of a set of campaigner nodes to mitigate a fake news process stemming from another set of nodes. It creates opportunities for designing a variety of objectives, e.g. minimizing the number of users who see fake news but were not reached by real news. We give the first derivation of second-order statistics of random exposure counts in the non-stationary case, which is essential in policy evaluation and improvement. By defining a

[1]School of Computational Science and Engineering, Georgia Tech. [2]Department of Mathematics and Statistics, Georgia State University. [3]School of Industrial and Systems Engineering, Georgia Tech.. Correspondence to: Mehrdad Farajtabar <mehrdad@gatech.edu>.

Proceedings of the 34^{th} International Conference on Machine Learning, Sydney, Australia, PMLR 70, 2017.

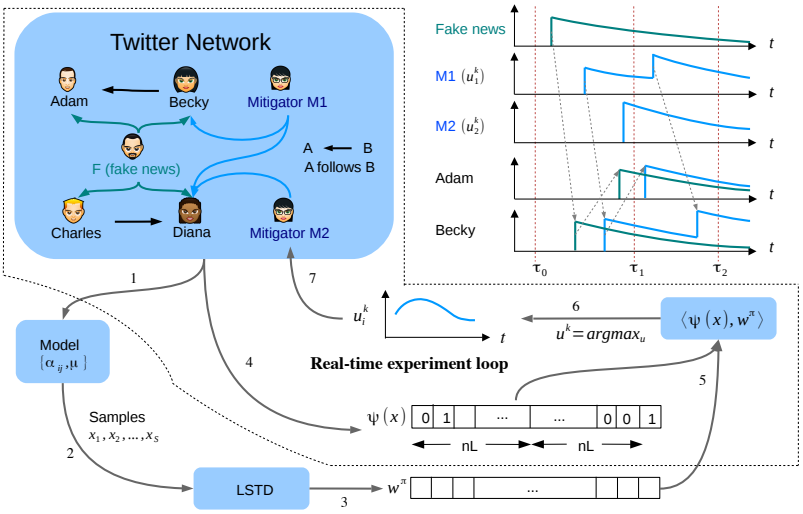


Figure 1. The framework of point process based intervention for countering fake news. (1-3) Offline learning of value function approximation weight vector using LSTD from transition samples generated from model. (4-7) Real-time intervention loop that uses feature representation of network state to choose optimal exogenous incentive for mitigator nodes.

state space for the network, formulating actions as exogenous intensity, and defining reward functions, we map the fake news mitigation problem to an optimal policy problem in a Markov decision process (MDP), which is solved by model-based least-squares temporal difference learning (LSTD) specific to the context of point processes. Furthermore, to the best of our knowledge, we are the first to conduct a real-time point process intervention experiment.

Related work. The emergence of social media as a prominent news source in the past few years raises concomitant concerns about the quality, truthfulness, and credibility of information presented (Mitra et al., 2017). To reduce the amount of labor-intensive manual fact-checking, there have been research efforts devoted to building classifiers to detect factuality of information, predicting credibility level of posts, and detecting controversial information from inquiry phrases (Mitra et al., 2017; Zeng et al., 2016; Zhao et al., 2015). These works mainly focused on extracting linguistic features from texts to determine the credibility of news and posts. Our focus in this paper, however, is to design an incentive strategy so that users can spontaneously take action to mitigate a real-world fake news epidemic.

Steering user activities by adding external incentives to the exogenous intensity of Hawkes processes was first considered in (Farajtabar et al., 2014). In (Farajtabar et al., 2016), a multistage campaigning method to optimally distribute incentive resources based on dynamic programming was developed. In these previous works, objective functions were designed using expected values of exposure counts rather than the stochastic exposure process, which may reduce the accuracy of solutions. Furthermore, it faced the demanding problem of computing the cost-to-go using the Hawkes model, while we address this using linear function approximation. For stationary Hawkes processes, second order statistics was derived in (Bacry & Muzy, 2014a;b); however, it is essential to compute both first and second order statistics for Hawkes processes in the *non-stationary* stages due to time sensitivity of the fake news mitigation task, and we derive it for the first time in this paper. Recent work has also applied methods in stochastic differential equations to the context of point processes, to find the best intensity for information guiding (Wang et al., 2016) and achieving highest visibility (Zarezade et al., 2017). While these works consider networks with only a single process, our work focuses on optimizing a mitigation process with respect to a second competing process. Finally, it's notable that our approach is related to but much more general than the influence maximization problems (Kempe et al., 2003; Bharathi et al., 2007) as it allows recurrent activity in social networks (in contrast to binary infection states), a variety of objectives (not only maximization), and budget constraints.

Reinforcement learning tackles the problem of finding good policies for actions to take in MDP where exact solutions are intractable, either due to size or lack of complete knowledge. Large-scale policy evaluation and iteration problems can be tackled by function approximation, which reduces the solution dimension using feature vector basis (Sutton & Barto, 1998). By adding control terms to a multivariate Hawkes process model of random network activities, fake news mitigation can be formulated as a policy optimization problem in an MDP. To address the randomness of Hawkes processes, batch reinforcement learning using samples collected from the trajectory of a fixed behavior policy can be applied (Antos et al., 2007). In particular, linear Least Squares Temporal Difference (LSTD) uses a batch of samples to learn a linear approximation of the value function under a policy with provable convergence (Bradtke & Barto, 1996). This *policy evaluation* step alternates with a model-based *policy improvement* step in a policy iteration to arrive at successively improved policies.

2. Preliminaries and Problem Statement

Multivariate Hawkes processes. Hawkes process is a doubly stochastic point process with self-excitations, meaning that past events increase the chance of arrivals of new events (Hawkes, 1971), and has been extensively used to model activities in social networks (Farajtabar et al., 2015; Linderman & Adams, 2014; He et al., 2015; Rizoiu et al.; Lee et al., 2016). Let t_ℓ be the time of the ℓ-th event, then the Hawkes process can be represented by the counting process $N(t) = \sum_{t_\ell \le t} h(t - t_\ell)$ that tracks the number of events up to time t, where $h(t)$ is the standard Heaviside function such that $h(t) = 1$ if $t \ge 0$ and $= 0$ if $t < 0$. The conditional intensity function of a point process is defined as the probability of observing an event in an infinitesimal window given the history. For Hawkes process it is given by $\lambda(t) = \mu + \sum_{t_\ell < t} \phi(t - t_\ell)$. Here, $\mu \ge 0$ is the exogenous (base) intensity and $\phi(t)$ is the Hawkes kernel that describes how fast the excitement of a past event decays. In this paper, we employ the standard (stationary) exponential Hawkes kernel, i.e., $\phi(t) = \alpha \mathrm{e}^{-\omega t} h(t)$ with $\omega > \alpha > 0$. In an n-dimensional multivariate Hawkes process (MHP), there are n such processes $N_1(t), \dots, N_n(t)$ that can also mutually excite one another, and the conditional intensity $\lambda(t) := \big(\lambda_1(t), \dots, \lambda_n(t)\big)^\top \in \mathbb{R}_+^n$ is given by $\lambda(t) = \mu + \int_0^t \Phi(t - s)\, \mathrm{d}N(s)$. Here, $N(t) := \big(N_1(t), \dots, N_n(t)\big)^\top \in \mathbb{N}_0^n$, $\mu := (\mu_1, \dots, \mu_n)^\top \in \mathbb{R}_+^n$, and $[\Phi(t)]_{ij} = \phi_{ij}(t) := \alpha_{ij} \mathrm{e}^{-\omega t} h(t)$. We let $\mathcal{H}(t)$ denote the filtration of $N(t)$, generated by the σ-algebra of history $\big\{(t_\ell, i_\ell) | t_\ell \le t\big\}$ of this point process, where $i_\ell \in \{1, \dots, n\}$ is the identity (node) of the ℓ-th event.

Network activities. We model the activities of both fake news and mitigation events as MHP in the network. Basically, MHP is a networked point process model with dependent dimensions (nodes), and can capture the underlying dynamics of networks and activities (Blundell et al., 2012; Xu et al., 2016; Guo et al., 2015). Define $F(t) = \big(F_1(t), \dots, F_n(t)\big)^\top \in \mathbb{N}_0^n$, where $F_i(t)$ counts the number of times user i shares a piece of news from the fake campaign up to time t. Similarly, define $M(t) = \big(M_1(t), \dots, M_n(t)\big)^\top \in \mathbb{N}_0^n$ for the mitigation process. Correspondingly, we have 2 intensity functions: $\lambda^M(t) = (\lambda_1^M(t), \dots, \lambda_n^M(t))^\top$ and $\lambda^F(t) = (\lambda_1^F(t), \dots, \lambda_n^F(t))^\top$ and two sets of exogenous intensities μ^M and μ^F.

Goal. Given that both $F(t)$ and $M(t)$ are modeled by the Hawkes processes, our goal is to find the optimal mitigation campaign by imposing interventions to users such that the mitigation effect (rigorously defined in sec. 3.1) can be maximized or equivalently the fake news be rectified under budget constraints. To this end, we measure the influence of fake news and mitigation activities using event exposures, describe the mechanism of mitigation interventions, and quantify the effect of interventions mathematically.

Event exposure. Event exposure is a quantitative measure of campaign influence, and is represented as a counting process, $\mathcal{E}(t) = \big(\mathcal{E}_1(t), \dots, \mathcal{E}_n(t)\big)^\top$. Here, $\mathcal{E}_i(t)$ records the number of times user i is exposed (she or one of her neighbors performs an activity) to a campaign $N(t)$ by time t. Let B be the adjacency matrix of the user network, i.e., $b_{ij} = 1$ if user i follows user j. Assume $b_{ii} = 1$ for all i. Then the exposure process is given by $\mathcal{E}(t) = BN(t)$. We define $\mathcal{F}(t) = BF(t)$ and $\mathcal{M}(t) = BM(t)$ as the fake news and mitigation processes, respectively.

Intervention. Suppose we can perform intervention by incentivizing a subset of users in the k-th stage during time $[\tau_k, \tau_{k+1})$ for $k = 0, 1, \dots$. For simplicity we consider uniform time duration $\tau_{k+1} - \tau_k = \Delta_T$ for all k, since generalization to nonuniform time durations is trivial. In order to steer the mitigation activities to counter the fake news (criteria given below) at these stages, we impose an additional constant intervention $u_i^k \ge 0$ to the exogenous intensity μ_i during time $[\tau_k, \tau_{k+1})$ for each stage $k = 0, 1, \dots$. The mitigation activity intensity at the k-th stage is $\lambda^M(t) = \mu + u^k + \int_0^t \Phi(t - s)\, \mathrm{d}M(s)$ for $t \in [\tau_k, \tau_{k+1})$. Note that the intervention itself exhibits a stochastic nature: adding u_i^k to μ_i is equivalent to incentivizing user i to increase her activity rate but it is still uncertain when she will perform an activity, which appropriately mimics the randomness in the real world.

Reward function. For each stage k, x^k (defined later) is the state of the whole MDP that encodes all the information from previous stages and u^k is the current control imposed at this stage. Let $\mathcal{M}_i^k(t; x^k, u^k) := \sum_j b_{ij} \int_{\tau_k}^t \mathrm{d}M_j(s)$ be the number of times user i is exposed to the mitigation campaign by time $t \in [\tau_k, \tau_{k+1})$ within stage k, then the goal is to steer the expected total number of exposure $\mathcal{M}_i^k(t; x^k, u^k)$ using u^k, *s.t.* the sum of reward functions $R(x^k, u^k)$ (rigorously defined in sec. 3.1) is maximized.

Problem statement. By observing the counting process in previous stages (summarized in a sequence of x^k) and taking the future uncertainty into account, the control problem is to design a policy π such that the controls $u^k = \pi(x^k)$ can maximize the total discounted objective $\mathbb{E}[\sum_{k=0}^\infty \gamma^k R^k]$, where $\gamma \in (0, 1]$ is the discount rate and R^k is the observed reward at stage k. In addition, we may have constraints on the amount of control, such as a budget constraint on the sum of all interventions to users at each stage, or a cap over the amount of intensity a user can handle. A feasible set or an action space over which we find the best intervention is represented as $U_k := \big\{u \in \mathbb{R}^n | u^\top c^k \le C_k, 0 \le u \le \alpha^k\big\}$. Here, c_i^k is the price per unit increase of exogenous intensity of user i and $C_k \in \mathbb{R}_+$ is the total budget at stage k. Also, α_i^k is the cap on the amount of activities of the user i.

3. Proposed Method

In this section, we present the formulation of reward functions in terms of event exposures of fake news and mitigation activities. Then we derive the key statistics of the MHP required for reward function evaluation, followed by the policy iteration scheme to find the optimal intervention.

3.1. Fake news mitigation

As we discussed above, the total reward of policy π is defined by the value function

$$V^{\pi}(x^0) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k R^k \,\middle|\, x^0\right] \tag{1}$$

for the initial state x^0 of fake and mitigation processes, where the observed reward R quantifies the effect of mitigation activities $M(t)$ in each stage and $\gamma \in (0, 1]$ is the discount rate. In this paper, we consider two types of reward functions $R(x, u)$:

1) *Correlation Maximization*: One possible way is to require correlation between mitigation exposures and fake news exposures: people exposed more to fake news should also be exposed more to the true news, to counter the fake news campaign. Therefore, we can form the reward function R in stage k as follows:

$$R(x^k, u^k) = \frac{1}{n} \mathcal{M}^k(\tau_{k+1}; x^k, u^k)^{\top} \mathcal{F}^k(\tau_{k+1}; x^k, u^k).$$

2) *Difference Minimization*: Suppose the goal is to minimize the number of unmitigated fake news events, then we can form a reward function R in stage k as the least squares of unmitigated numbers:

$$R(x^k, u^k) = \frac{-1}{n} \left\| \mathcal{M}^k(\tau_{k+1}; x^k, u^k) - \mathcal{F}^k(\tau_{k+1}; x^k, u^k) \right\|^2$$

These are two sample realizations of the MHP-MDP based intervention one can formulate, among many others. To solve the policy optimization problem $\operatorname{argmax}_{\pi} V^{\pi}(x^0)$ for V^{π} defined in (1), we need to evaluate the value function V^{π} for any given policy π, which requires the first and second order statistics (moments) of any multivariate Hawkes processes $N(t)$, as we derive next.

3.2. Second order statistics of non-stationary MHP

For an n-dim MHP $N(t)$ with standard exponential kernel $\Phi(t)$, the following proposition provides closed-form solution of the mean intensity $\eta(t) := \mathbb{E}[\lambda(t)]$ for both constant and time-varying exogenous intensity $\mu(t)$:

Proposition 1 (Theorem 3 (Farajtabar et al., 2014; 2016))**.** *Let $N(t)$ be an n-dimensional MHP defined in sec. 2 with exogenous intensity $\mu(t)$ and Hawkes kernel $\Phi(t) = A\mathrm{e}^{-\omega t}h(t)$, then the mean intensity $\eta(t)$ is given by*

$$\eta(t) = \left[\mathrm{e}^{(A-\omega I)t} + \omega(A-\omega I)^{-1}\left(\mathrm{e}^{(A-\omega I)t} - I\right)\right]\mu(t). \tag{2}$$

Let $\Lambda(t) = \int_0^t \lambda(s)\,\mathrm{d}s$ be the compensator of $N(t)$, then by Doob-Meyer's decomposition theorem $N(t) - \Lambda(t)$ is a zero mean martingale. This implies that the first order statistics $\mathbb{E}[N(t)]$ can be obtained by $\mathbb{E}[N(t)] = \mathbb{E}[\Lambda(t)] = \mathbb{E}[\int_0^t \lambda(s)\,\mathrm{d}s] = \int_0^t \mathbb{E}[\lambda(s)]\,\mathrm{d}s = \int_0^t \eta(s)\,\mathrm{d}s$ using eq. (2).

To evaluate the reward function R defined previously, we need to derive second order statistics of multivariate Hawkes process $N(t)$ in its non-stationary stage. The following theorem states the key ingredients for the second order statistics. The proof is provided in the appendix.

Theorem 2. *Let $N(t)$ be an n-dim MHP with exogenous intensity μ and Hawkes kernel Φ defined in sec. 2, then the second order statistics of $N(t)$ for $t, t' \geq 0$ is given by*

$$\begin{aligned}\mathbb{E}\left[\mathrm{d}N(t)\,\mathrm{d}N(t')^{\top}\right] &= G(t', t)^{\top}\Sigma(t')\,\mathrm{d}t\,\mathrm{d}t' + \\ &\quad \delta(t-t')\Sigma(t')\,\mathrm{d}t\,\mathrm{d}t' + \eta(t)\eta(t')^{\top}\,\mathrm{d}t\,\mathrm{d}t'\end{aligned} \tag{3}$$

where $\eta(t) = \mathbb{E}[\lambda(t)]$ is given in (2)*, $\Sigma(t) = \mathrm{diag}([\eta_i(t)])$ is diagonal, and G is the unique solution of*

$$G(t', t) = G(t', t) * \Phi(t) + \Phi(t - t') - \delta(t - t')I. \tag{4}$$

Moreover $G(t', t)^{\top}\Sigma(t') = \Sigma(t)G(t, t')$ for all $t, t' \geq 0$.

Based on Theorem 2, we can compute second order statistics such as $\mathbb{E}[N_i(t)N_j(t')]$ for all i, j and $t, t' \geq 0$.

3.3. State Representation

Hawkes process is non-Markovian and one needs complete knowledge of the history to characterize the entire process. However, when the standard exponential kernel $\Phi(t, s) = Ae^{-\omega(t-s)}h(t-s)$ is employed, the effect of history up to time τ_k on the future $t > \tau_k$ can be cleverly summarized by one scalar per dimension (Simma & Jordan, 2012; Farajtabar et al., 2016). For $1 \leq i \leq n$, define $y_i^k := \lambda_i^{k-1}(\tau_k) - u_i^{k-1} - \mu_i$, (and $y_0^i = 0$ by convention), then the intensity due to events of all previous k stages can be written as $\int_0^{\tau_k} Ae^{-\omega(t-s)}\,\mathrm{d}N(s) = y^k e^{-\omega(t-\tau_k)}$. In other words, y^k is sufficient to encode the information of activities in the past k stages that are relevant to future. Note that we have two separate y_M^k and y_F^k to track the dynamics of both mitigation and fake processes.

Also, in order to tackle objectives over multiple stages, we add aggregated number of events at L previous Δ_f-time intervals over all dimensions. Define a vector $z^k \in \mathbb{R}^{nL}$ where $z^k_{(l-1)n+i} = \int_{\tau_k - l\Delta_f}^{\tau_k - (l-1)\Delta_f} \mathrm{d}N_i(s)$ for $1 \leq i \leq n$ and $1 \leq l \leq L$. In other words, $z^k_{(l-1)n+i}$ records the number of events of i-th dimension in the l-th interval of length Δ_f prior to time τ_k. For example, choosing $\Delta_f = \Delta_T$ and setting $L = 2$ means that events from the two most recent stages are counted. Similarly, we have two separate z_M^k and z_F^k corresponding to the two processes. Now, the state vector $x^k \in \mathbb{R}^{2nL+2n}$ is the concatenation of the above four vectors $x^k = [y_M^k; y_F^k; z_M^k; z_F^k]$.

Algorithm 1 LSTD policy iteration in point processes

Input: set of samples $\mathcal{S}$, feature $\psi(\cdot)$, discount γ
repeat
  Initialize $A^\pi = 0$ and $b^\pi = 0$.
  for each state $x \in \mathcal{S}$ **do**
    $A^\pi \leftarrow A^\pi + \psi(x)(\psi(x) - \gamma\psi(x'))^\top$
    $b^\pi \leftarrow b^\pi + \psi(x) r^\pi$
  end for
  $w^\pi \leftarrow (A^\pi)^{-1} b^\pi$
  for each state $x \in \mathcal{S}$ **do**
    $\pi(x) \leftarrow \operatorname{argmax}_u \{\mathbb{E}[R(x,u)] + \gamma \mathbb{E}[V^\pi(x')|u, w^\pi]\}$
  end for
until $\|\Delta w^\pi\| < 0.1$
return w^π

3.4. Least Squares Temporal Difference

The optimal value function satisfies the Bellman equation:

$$V^\pi(x) = \mathbb{E}[R(x, \pi(x))] + \gamma \mathbb{E}[V^\pi(x')], \tag{5}$$

where x' is the next state after taking action based on policy π at state x. Least squares temporal difference learning (LSTD) is a sample-efficient procedure for policy evaluation, which subsequently facilitates policy improvement. The value function is approximated by $\hat{V}^\pi(x) = \sum_{d=1}^{D} w_d^\pi \psi_d(x)$, where ψ_d is the d-th feature of state x and w_d^π is its coefficient for policy π. This can be compactly represented as $\hat{V}^\pi(x) = \psi(x)^\top w^\pi$, where $\psi(x) = (\psi_1(x), \ldots, \psi_D(x))^\top$. The following presents our choice of features, and explains the policy evaluation and improvement steps inspired by LSTD(0) (Sutton & Barto, 1998).

Features. The number of events in a few recent consecutive intervals of point processes have been used as a reliable feature to parameterize point processes (Parikh et al., 2012; Qin & Shelton, 2015; Lian et al., 2015). Following their work we take L prior intervals of length Δ_f for each dimension of the fake news process and record the number of events in that period as one feature. $\psi^k_{(l-1)n+i} = z^k_{(l-1)n+i}$ for $1 \le i \le n$ and $1 \le l \le L$. This will count for nL features. Similarly we take nL features from the mitigation process. Finally, we add a last feature $\psi^k_{2nL+1} = 1$ as the bias term. Therefore, $\psi^k = [z^k_M; z^k_F; 1]$ and the feature space has dimension $D = 2nL + 1$.

Policy Evaluation. Substituting the approximation into the Bellman equation, we have:

$$\psi(x)^\top w^\pi = \mathbb{E}[R(x, \pi(x))] + \gamma \mathbb{E}[\psi(x')^\top] w^\pi. \tag{6}$$

To find the best fit of w^π we have to consider all possible x; however, since the state space is infinite-dimensional, enumerating all states is impossible and we utilize a set $\mathcal{S}$ of samples $\mathcal{S} = \{x_1, \ldots, x_S\}$.

Algorithm 2 Real-time fake news mitigation

Input: network A, learned w^π, feature $\psi(\cdot)$, discount γ
repeat
  Observe state x of the network activities
  $u = \operatorname{argmax}_a \{\mathbb{E}[R(x,a)] + \gamma \mathbb{E}[V^\pi(x')|a, w^\pi]\}$
  Add u to base exogenous intensity μ and generate mitigation event times $\{t_i\}$ using point process model
  Create posts at times $\{t_i\}$ using campaigner accounts
until end of campaign

Let $\psi(x_s) = \psi_s \in \mathbb{R}^D$, $\mathbb{E}[\psi(x'_s)] = \psi'_s \in \mathbb{R}^D$, and $r^\pi_s = \mathbb{E}[R(x_s, \pi(x_s))] \in \mathbb{R}$. Then define matrices of current features $\Psi = [\psi_1^\top; \ldots; \psi_S^\top]^\top \in \mathbb{R}^{S \times D}$ and next features $\Psi' = [{\psi'}_1^\top; \ldots; {\psi'}_S^\top]^\top \in \mathbb{R}^{S \times D}$, the rewards $r^\pi = [r_1^\pi, \ldots, r_S^\pi]^\top \in \mathbb{R}^S$, and the sample value functions as $v^\pi = [V^\pi(x_1), \ldots, V^\pi(x_S)]^\top \in \mathbb{R}^S$. Appendix C presents how we leverage the first and second order statistics of Hawkes process to find $\mathbb{E}[R(x,u)]$ and $\mathbb{E}[V^\pi(x')]$. Given the above definition, the Bellman optimality of eq. (6) can be written in matrix format:

$$v^\pi = \Psi w^\pi = r^\pi + \gamma \Psi' w^\pi \triangleq T^\pi v^\pi, \tag{7}$$

where T^π is the Bellman optimality operator. A way to find a good estimate is to force the approximate value function to be a fixed point of the optimality equation under the Bellman operator, *i.e.*, $T^\pi \hat{v}^\pi \approx \hat{v}^\pi$. (Lagoudakis & Parr, 2003). For that, the fixed point has to lie in the space of approximate value functions, spanned by the basis functions Ψ. $\hat{v}^\pi$ lies in that space by definition, but $T^\pi \hat{v}^\pi$ may have an orthogonal component and must be projected. This is achieved by the orthogonal projection operator $(\Psi(\Psi^\top \Psi)^{-1} \Psi^\top)$. Therefore the approximate value function $\hat{v}^\pi$ must be invariant under one application of the Bellman operator T^π followed by orthogonal projection:

$$\hat{v}^\pi = \Psi(\Psi^\top \Psi)^{-1} \Psi^\top (T^\pi \hat{v}^\pi). \tag{8}$$

By substituting the linear approximation $\Psi w^\pi = v^\pi$ into the above equation and some manipulations, we get a $D \times D$ linear systems of equations $A^\pi \omega^\pi = b^\pi$, where $A^\pi = \Psi^\top(\Psi - \gamma \Psi')$ and $b^\pi = \Psi^\top r^\pi$, and whose solution is the fitted coefficients w^π. It has been shown that the estimated w^π converges to the best w^* as the available number of samples tends to infinity (Bradtke & Barto, 1996). Appendix B presents a detailed derivation.

Policy Improvement. The second part of the algorithm implements policy improvement, i.e., getting an improved policy π' via one-step look-ahead as follows:

$$\pi'(x) = \operatorname*{argmax}_u \mathbb{E}[R(x,u) + \gamma V^\pi(x')]. \tag{9}$$

LSTD(0) alternates between the policy improvement and policy evaluation iteratively until w^π converges (Bradtke & Barto, 1996). Alg. 1 summarizes this procedure.

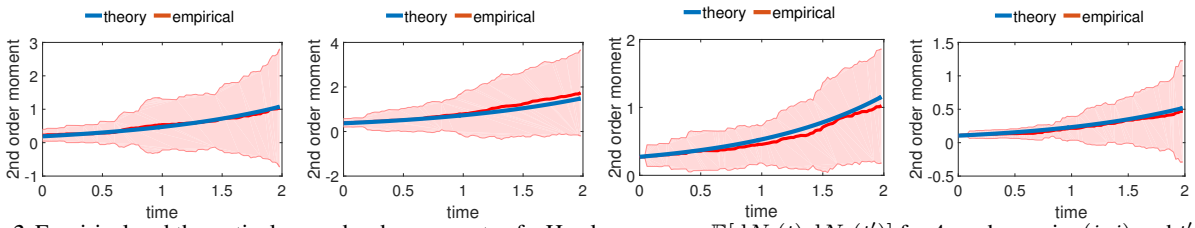


Figure 2. Empirical and theoretical second order moments of a Hawkes process, $\mathbb{E}[\mathrm{d}N_i(t)\,\mathrm{d}N_j(t')]$ for 4 random pairs (i, j) and $t' = 0$ and varying t from 0 to 2.

LSTD in Hawkes context. LSTD is particularly suitable to the problem we are interested in. It learns the value function $V^\pi(x)$, and as such, policy improvement can be challenging without knowing the model. Because of this, methods that aim to learn the Q-function $Q^\pi(x, u)$, such as LSPI (Lagoudakis & Parr, 2003), are widely applied. The downside of Q-function based methods is that they typically require more samples than learning the value function. Yet, in our setup, learning the value function is sufficient, by writing the action-value function as $Q^\pi(x, u) = \mathbb{E}[R(x, u) + V^\pi(x')]$, and observing that the learned model of the multivariate Hawkes process enables analytical computation of the expectation (see Appendix C for details):

$$\begin{aligned}
&\mathbb{E}[V^\pi(x')] \\
&= \sum_{i=1}^{n}\sum_{l=1}^{L-1} w^\pi_{ln+i} z^{k-1}_{M,(l-1)n+i} + w^\pi_{nL+ln+i} z^{k-1}_{F,(l-1)n+i} \\
&+ \sum_{i=1}^{n} w_i \mathbb{E}[z^k_{M,i}] + w_{nL+i}\mathbb{E}[z^k_{F,i}] + w^\pi_{2nL+1}, \\
&\mathbb{E}[R(x, u)] = \frac{1}{n}\mathbb{E}[z^k_M]^\top B^\top B\, \mathbb{E}[z^k_F], \qquad \% \text{ correlation} \\
&\mathbb{E}[R(x, u)] = -\frac{1}{n}\mathbb{E}[z^k_M{}^\top B^\top B\, z^k_M] - \frac{1}{n}\mathbb{E}[z^k_F{}^\top B^\top B\, z^k_F] \\
&+ \frac{2}{n}\mathbb{E}[z^k_M]^\top B^\top B\, \mathbb{E}[z^k_F]. \qquad \% \text{ difference}
\end{aligned}$$

We require much fewer samples to learn $V^\pi(x)$ compared to learning an approximate $Q^\pi(x, u)$, and in particular compared to LSPI we avoid explicitly discretizing the continuous action space from which the action u is chosen.

We further remark that the policy improvement step finds the optimal action u at any state x by computing $\mathrm{argmax}_u\, \mathbb{E}[R(x, u) + V^\pi(x')]$, where the action u to be optimized appears in the calculation of both the expected current reward and the expected value at the next state. This optimization problem is convex under our choice of reward functions and the form of the Hawkes conditional intensity.

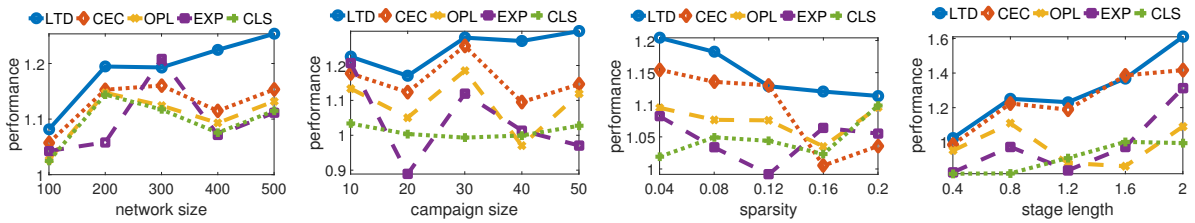
After learning the optimal policy (implicit by w^π of the linearly-approximated value function) we start at the real-time intervention part. By observing the state we find the optimal intervention intensity by simply solving eq. (9). Alg. 2 summarizes the real-time mitigation procedure.

4. Experiments

We evaluate our fake news mitigation framework by both simulated and real-time real-world experiments and show our approach, Least-squares Temporal Difference (LTD), significantly outperforms several state-of-the-art methods and alternatives: CEC (an approximate dynamic programming), OPL (an open loop optimization), CLS (a centrality based measure), EXP (an exposure based centrality measure), and RND (the random policy). Their details are given in appendix D. Before explaining the intervention results we verify the theoretical second order statistics in Fig. 2 whose details can be found in appendix E. Furthermore, we examine convergence properties and representative power of linear features in Fig. 5 with the details postponed to appendix F due to space limitations.

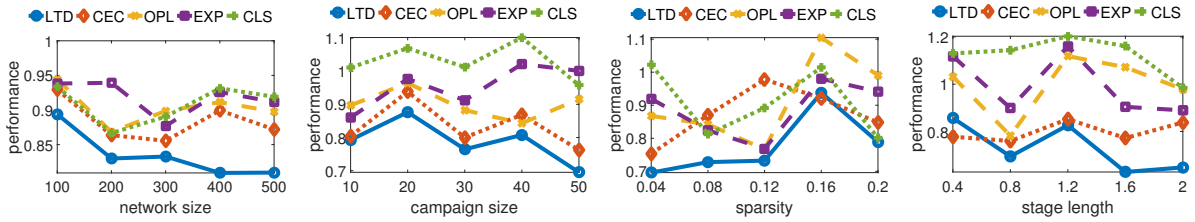
4.1. Synthetic Experiments

Setup. For all except the experiment over network size, the networks were generated synthetically with $n = 300$ nodes. Endogenous intensity coefficients were set as $a_{ij} \sim \mathcal{U}[0, 0.5]$. To mimic real world networks, sparsity was set to 0.02, *i.e.*, each edge was kept with probability 0.02. The influence matrix was scaled appropriately such that the spectral radius is a random number smaller than one to ensure the stability of the process. The Hawkes kernel parameter was set to $\omega = 1$, which means loosing roughly 63 % of influence after 1 unit of time (minutes, hours, etc). Both fake news and mitigation processes obey these network settings. Among n nodes, we assume 20 nodes create fake news and another 20 nodes can be incentivized (via the exogenous intensity) to spread true news. Each stage has length of $\Delta_T = 1$. The discount factor was set to $\gamma = 0.7$. For determining features, we set $L = 2$ and we choose $\Delta_f = \Delta_T$ for simplicity. The upper bound for the intervention intensity was chosen by $\alpha_i \sim \mathcal{U}[0, 0.5]$. The price of each person was $c^k_i = 1$, and the total budget at stage k was randomly generated as $C_k \sim (n \times \mathcal{U}[0, 0.5])$. 1000 randomly sampled states were used for the LSTD algorithm. To evaluate a policy (learnt by an algorithm) we simulated the network under that policy 50 times and take the discounted total reward averaged over these 50 runs as an empirical valuation of the policy. Furthermore, each single run was simulated for 10 consecutive stages; from the eleventh stage onward, the objectives contribute 0.02 of the total reward and can be safely discarded. For all experiments, the above settings are assumed unless it is explicitly mentioned otherwise.



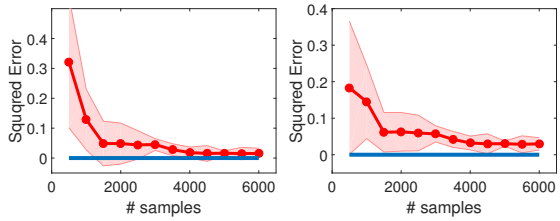
(a) (b) (c) (d)

Figure 3. Performance improvement of different methods over the random policy on synthetic networks for correlation maximization



(a) (b) (c) (d)

Figure 4. Performance improvement of different methods over the random policy on synthetic networks for distance minimization



(a) Correlation (b) Difference

Figure 5. Convergence of linear approximated value function

Intervention results. Fig. 3 demonstrates the performance of different methods. Performance of a policy is quantified as the ratio of the total reward achieved by running the policy, over the total reward achieved by the random policy (RND). This allows us to compare the effectiveness of the algorithms over a variety of settings. All the results reported are averages over 10 runs with random networks generated according to the above setup. Overall, it is clear that LTD is almost consistently the best. It improves over the random policy by roughly 20 percent. CEC is the second best and shows the effectiveness of multi-stage and closed loop intervention. This validates our intuition that although CEC computes the reward from both fake news and mitigation processes, the lack of explicit features corresponding to previous events in its value function prevents it from learning the reason for the reward. Roughly, OPL is the third best algorithm, due to its negligence of the state and the actual events that occurred. Next, comes the EXP algorithm followed by the CLS. The poor performance of these (compared to others) shows that structural properties are not sufficient to tackle the fake news mitigation problem. EXP is roughly better than CEC because it heuristically takes into account the fake news exposure.

Fig. 3-a shows the performance with respect to increasing network size. The difference between alternative methods and the gap between LTD and others increase with the network size. Furthermore, the performance of all methods show an increase over random policy when the problem size gets larger. This illustrates the fact that efficient distribution of budget matters more when confronted with problems of increasing complexity and size.

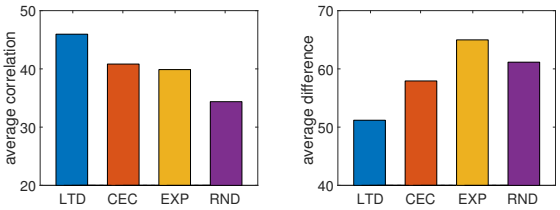
Fig. 3-b shows the performance with respect to increasing the mitigation campaign size. Larger campaigns imply greater flexibility of intervention, which can be exploited by clever algorithms to achieve higher performance.

Fig. 3-c shows the performance with respect to increasing sparsity of the network. Interestingly, the performance of all the algorithms move towards to the random policy as the network becomes denser. This can be understood by considering a complete graph, so that no matter how and to whom we distribute the mitigation budget, all the nodes are exposed to the mitigation campaign almost equally. However, since real social networks are usually sparse, the effectiveness of the proposed method stands out.

Finally, Fig. 3-d shows the performance with respect to the length of an stage. Longer stage lengths increase the potential for a good policy to attain higher reward than a random policy, and this is reflected by the sharp increase and larger performance gap between LTD and others for longer lengths. We observe the same patterns for the distance minimization in Fig. 4 problem and avoid repeating them.

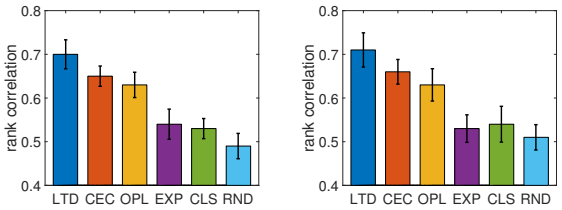
4.2. Real experiments

In this section we explain our real-time intervention results. To the best of our knowledge, we are the first to employ a real-time experiment to evaluate a point process based social network intervention strategy.



(a) Correlation (b) Difference

Figure 6. Results of fake news mitigation on Twitter network



(a) Correlation (b) Difference

Figure 7. Rank correlation for prediction

Setup. Using five Twitter accounts, each of which made five posts on machine learning topics at random times per day for a span of two months (Nov.-Dec. 2016), we accumulated a network of 1894 real users with 23407 directed edges in total. We used this historical data to learn the network parameters $\{\alpha_{ij}, \mu^i\}$ using maximum likelihood (similar to related work (Zhou et al., 2013; Farajtabar et al., 2014)) with one hour as the time resolution and the kernel decay parameter ω set to 0.1. As illustrated in Fig.1 the optimal policy was learned using LSTD and policy improvement. Then the real-time experiment starts: Two of the accounts, interpreted as the source of fake news, continued to behave using the same randomized policy as they did in the data collection stage, while the posting times of the other three accounts were generated from $(u_1, u_2, u_3)^T$, produced by our LTD strategy or a competitor strategy. Each policy was run for 10 stages of length 12 hours. Therefore, $\Delta_T = \Delta_f = 12$. Since both fake news and mitigation accounts were tweeting random posts on machine learning, we assume negligible bias in the content that can confound the performance. At the end of each stage, all retweets–by users within the network–of the posts made during the two most recent stages were used to construct the feature vector and compute the value function, which was used to find the optimal intervention for the next stage. The methods CEC and OPL belong to the same category, and it has been shown that CEC outperforms OPL in (Farajtabar et al., 2016). Furthermore, EXP and CLS also belong to similar families and our synthetic experiments confirm the superiority of the former. So, to save time in real interventions, we only test CEC from the first and EXP from the second pair, and compare them with the random policy (RND) and with our algorithm (LTD).

Real-time intervention results. Fig. 6 shows the performance of our results compared to competitors. The results show that our approach outperforms the other three baselines by a reasonable margin. As expected CEC is the second best algorithm with a margin of 5 for the correlation maximization objective. It translates to increase in amount of correlation equal to 5, which is a noticeable amount. Furthermore, in the difference minimization task, our approach reached around 7 in difference. This means that we decreased the difference in exposure to the two processes to less 2.6 per user, which is considerable improvement. For both tasks, LTD made more mitigation posts over all daytime phases than it did over all nighttime phases, whereas the competitor strategies did the opposite. This could be a reason for its better performance. One surprising fact is that the number of retweets by users outside the network, which was not used for our features, can exceed the number of retweets by users within the network. This is because the "hashtag" feature on Twitter allows posts to be seen by a much larger set of users, who do not necessarily follow the source accounts. In addition to retweets, users can also "like" a post, indicating that they were exposed to fake or real news; while we measured this, we did not include it in the reward. Future experiments can use these two observations to widen the experimental scope and more accurately measure the effectiveness of a mitigation strategy. Despite having these limitations, our experiment serves as a proof-of-concept for the applicability of point process based intervention in networks, and–to the best of our knowledge–is the first to verify the superiority of a method in a real-time, real-world intervention setting.

Prediction evaluation results. The previous part describes the more interesting evaluation scheme of real-time intervention in a social media platform. In this part, we used historical real data to mimic this procedure. We extracted 12 full 10-stage trajectory of events from the 2-month random policy historical data. For any of these 10 pairs, the methods were evaluated according to how well they predict the relative ordering among these 12 trajectories (with respect to the objective function). To evaluate each method, we created a sorted list of these 12 trajectories according to increasing objective, and created a second list sorted by increasing closeness to the intervention method. This closeness is the mean squared error between the prescribed intervention and actual intensity, which we inferred using maximum likelihood. Then, by computing the rank correlation of the two sorted lists, and repeating for each of the five methods, we can find out how well they perform on the prediction task. A better predictor is expected to be a better mitigation strategy. Fig. 7 shows the performance.

A short discussion is presented in appendix G.

Acknowledgement. This project is supported in part by NSF IIS-1639792, CNS-1409635, NSF DMS-1620342, NSF IIS-1218749, NIH BIGDATA 1R01GM108341, NSF CAREER IIS-1350983, ONR N00014-15-1-2340, NVIDIA, Intel, and Amazon AWS.

References

Allcott, Hunt and Gentzkow, Matthew. Social media and fake news in the 2016 election. Technical report, National Bureau of Economic Research, 2017.

Antos, András, Szepesvári, Csaba, and Munos, Rémi. Value-iteration based fitted policy iteration: learning with a single trajectory. In *Approximate Dynamic Programming and Reinforcement Learning, 2007. ADPRL 2007. IEEE International Symposium on*, pp. 330–337. IEEE, 2007.

Bacry, Emmanuel and Muzy, Jean-François. Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166, 2014a.

Bacry, Emmanuel and Muzy, Jean-Francois. Second order statistics characterization of hawkes processes and non-parametric estimation. *arXiv preprint arXiv:1401.0903*, 2014b.

Bertsekas, Dimitri P. *Dynamic programming and optimal control*, volume 1. 1995.

Bharathi, Shishir, Kempe, David, and Salek, Mahyar. Competitive influence maximization in social networks. In *International Workshop on Web and Internet Economics*, pp. 306–311. Springer, 2007.

Blundell, Charles, Beck, Jeff, and Heller, Katherine A. Modelling reciprocating relationships with hawkes processes. In *Advances in Neural Information Processing Systems*, pp. 2600–2608, 2012.

Bradtke, Steven J. and Barto, Andrew G. Linear least-squares algorithms for temporal difference learning. *Machine Learning*, 22(1):33–57, 1996. ISSN 1573-0565. doi: 10.1007/BF00114723. URL http://dx.doi.org/10.1007/BF00114723.

Chen, Duanbing, Lü, Linyuan, Shang, Ming-Sheng, Zhang, Yi-Cheng, and Zhou, Tao. Identifying influential nodes in complex networks. *Physica a: Statistical mechanics and its applications*, 391(4):1777–1787, 2012.

De Arruda, Guilherme Ferraz, Barbieri, André Luiz, Rodríguez, Pablo Martín, Rodrigues, Francisco A, Moreno, Yamir, and da Fontoura Costa, Luciano. Role of centrality for the identification of influential spreaders in complex networks. *Physical Review E*, 90(3):032812, 2014.

Farajtabar, Mehrdad, Du, Nan, Gomez-Rodriguez, Manuel, Valera, Isabel, Zha, Hongyuan, and Song, Le. Shaping social activity by incentivizing users. In *Advances in neural information processing systems*, pp. 2474–2482, 2014.

Farajtabar, Mehrdad, Wang, Yichen, Rodriguez, Manuel Gomez, Li, Shuang, Zha, Hongyuan, and Song, Le. Coevolve: A joint point process model for information diffusion and network co-evolution. In *Advances in Neural Information Processing Systems*, pp. 1954–1962, 2015.

Farajtabar, Mehrdad, Ye, Xiaojing, Harati, Sahar, Song, Le, and Zha, Hongyuan. Multistage campaigning in social networks. In *Advances in Neural Information Processing Systems*, pp. 4718–4726, 2016.

Gao, Zhenxiang, Shi, Yan, and Chen, Shanzhi. Identifying influential nodes for efficient routing in opportunistic networks. *Journal of Communications*, 10(1), 2015.

Gottfried, Jeffrey and Shearer, Elisa. *News Use Across Social Media Platforms 2016*. Pew Research Center, May 2016.

Guo, Fangjian, Blundell, Charles, Wallach, Hanna, and Heller, Katherine. The bayesian echo chamber: Modeling social influence via linguistic accommodation. In *Artificial Intelligence and Statistics*, pp. 315–323, 2015.

Hawkes, Alan G. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

He, Xinran, Rekatsinas, Theodoros, Foulds, James, Getoor, Lise, and Liu, Yan. Hawkestopic: A joint model for network inference and topic modeling from text-based cascades. In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pp. 871–880, 2015.

Kempe, David, Kleinberg, Jon, and Tardos, Éva. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137–146. ACM, 2003.

Lagoudakis, Michail G and Parr, Ronald. Least-squares policy iteration. *Journal of machine learning research*, 4(Dec):1107–1149, 2003.

Lee, Young, Lim, Kar Wai, and Ong, Cheng Soon. Hawkes processes with stochastic excitations. In *Proceedings of The 33rd International Conference on Machine Learning*, pp. 79–88, 2016.

Lian, Wenzhao, Henao, Ricardo, Rao, Vinayak, Lucas, Joseph E, and Carin, Lawrence. A multitask point process predictive model. In *ICML*, pp. 2030–2038, 2015.

Linderman, Scott W and Adams, Ryan P. Discovering latent network structure in point process data. In *ICML*, pp. 1413–1421, 2014.

Mitra, Tanushree, Wright, G, and Gilbert, Eric. A parsimonious language model of social media credibility across disparate events. In *Proc. CSCW*, 2017.

Mosseri, Adam. Addressing hoaxes and fake news, 2016. URL http://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/.

Parikh, Ankur P, Gunawardana, Asela, and Meek, Christopher. Conjoint modeling of temporal dependencies in event streams. *BMAW-12 Preface*, 2012.

Qin, Zhen and Shelton, Christian R. Auxiliary gibbs sampling for inference in piecewise-constant conditional intensity models. In *UAI*, pp. 722–731, 2015.

Rizoiu, Marian-Andrei, Xie, Lexing, Sanner, Scott, Cebrian, Manuel, Yu, Honglin, and Van Henteryck, Pascal. Expecting to be hip: Hawkes intensity processes for social media popularity.

Silverman, Craig. This analysis shows how viral fake election news stories outperformed real news on facebook, 2016. URL https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook.

Simma, Aleksandr and Jordan, Michael I. Modeling events with cascades of poisson processes. *arXiv preprint arXiv:1203.3516*, 2012.

Sutton, Richard S. and Barto, Andrew G. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1st edition, 1998. ISBN 0262193981.

Twitter. Types of tweets and where they appear, 2016. URL https://support.twitter.com/articles/119138.

Wang, Yichen, Theodorou, Evangelos, Verma, Apurv, and Song, Le. Steering opinion dynamics in information diffusion networks. *CoRR*, abs/1603.09021, 2016. URL http://arxiv.org/abs/1603.09021.

Xu, Qiaofeng, Yang, Deshan, Tan, Jun, Sawatzky, Alex, and Anastasio, Mark A. Accelerated fast iterative shrinkage thresholding algorithms for sparsity-regularized cone-beam ct image reconstruction. *Medical Physics*, 43(4):1849–1872, 2016.

Zarezade, A., Upadhyay, U., Rabiee, H., and Gomez-Rodriguez, M. Redqueen: An online algorithm for smart broadcasting in social networks. In *WSDM '17: Proceedings of the 10th ACM International Conference on Web Search and Data Mining*, 2017.

Zeng, Li, Starbird, Kate, and Spiro, Emma S. # unconfirmed: Classifying rumor stance in crisis-related social media messages. In *Tenth International AAAI Conference on Web and Social Media*, 2016.

Zhao, Zhe, Resnick, Paul, and Mei, Qiaozhu. Enquiring minds: Early detection of rumors in social media from enquiry posts. In *Proceedings of the 24th International Conference on World Wide Web*, pp. 1395–1405. ACM, 2015.

Zhou, Ke, Zha, Hongyuan, and Song, Le. Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, pp. 641–649, 2013.