Multistage Campaigning in Social Networks

Mehrdad Farajtabar[*] **Xiaojing Ye**[◇] **Sahar Harati**[†]

Le Song[*] **Hongyuan Zha**[*]

Georgia Institute of Technology[*] Georgia State University[◇] Emory University[†]

mehrdad@gatech.edu xye@gsu.edu sahar.harati@emory.edu

{lsong,zha}@cc.gatech.edu

Abstract

We consider the problem of how to optimize multi-stage campaigning over social networks. The dynamic programming framework is employed to balance the high present reward and large penalty on low future outcome in the presence of extensive uncertainties. In particular, we establish theoretical foundations of optimal campaigning over social networks where the user activities are modeled as a multivariate Hawkes process, and we derive a time dependent linear relation between the intensity of exogenous events and several commonly used objective functions of campaigning. We further develop a convex dynamic programming framework for determining the optimal intervention policy that prescribes the required level of external drive at each stage for the desired campaigning result. Experiments on both synthetic data and the real-world MemeTracker dataset show that our algorithm can steer the user activities for optimal campaigning much more accurately than baselines.

1 Introduction

Obama was the first US president in history who successfully leveraged online social media in presidential campaigning, which has been popularized and become a ubiquitous approach to electoral politics (such as in the on-going 2016 US presidential election) in contrast to the decreasing relevance of traditional media such as TV and newspapers [1, 2]. The power of campaigning via social media in modern politics is a consequence of online social networking being an important part of people's regular daily social lives. It has been quite common that individuals use social network sites to share their ideas and comment on other people's opinions. In recent years, large organizations, such as governments, public media, and business corporations, also start to announce news, spread ideas, and/or post advertisements in order to steer the public opinion through social media platform. There has been extensive interest for these entities to influence the public's view and manipulate the trend by incentivizing influential users to endorse their ideas/merits/opinions at certain monetary expenses or credits. To obtain most cost-effective trend manipulations, one needs to design an optimal campaigning strategy or policy such that quantities of interests, such as influence of opinions, exposure of a campaign, adoption of new products, can be maximized or steered towards the target amount given realistic budget constraints.

The key factor differentiating social networks from traditional media is *peer influence*. In fact, events in an online social network can be categorized roughly into two types: endogenous events where users just respond to the actions of their neighbors within the network, and exogenous events where users take actions due to drives external to the network. Then it is natural to raise the following fundamental questions regarding optimal campaigning over social networks: can we model and exploit those event data to steer the online community to a desired exposure level? More specifically, can we drive the overall exposure to a campaign to a certain level (e.g., at least twice per week per user) by incentivizing a small number of users to take more initiatives? What about maximizing the overall exposure for a target group of people?

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

More importantly, those exposure shaping tasks are more effective when the interventions are implemented in multiple stages. Due to the inherent uncertainty in social behavior, the outcome of each intervention may not be fully predictable but can be anticipated to some extent before the next intervention happens. A key aspect of such situations is that interventions can't be viewed in isolation since one must balance the desire for high present reward with the penalty of low future outcome.

In this paper, the *dynamic programming* framework [3] is employed to tackle the aforementioned issues. In particular, we first establish the fundamental theory of optimal campaigning over social networks where the user activities are modeled as a multivariate Hawkes process (MHP) [4, 5] since MHP can capture both endogenous and exogenous event intensities. We also derive a time dependent linear relation between the intensity of exogenous events and the overall exposure to the campaign. Exploiting this connection, we develop a convex dynamic programming framework for determining the optimal intervention policy that prescribes the required level of external drive at each stage in order for the campaign to reach a desired exposure profile. We propose several objective functions that are commonly considered as campaigning criteria in social networks. Experiments on both synthetic data and real world network of news websites in the MemeTracker dataset show that our algorithms can shape the exposure of campaigns much more accurately than baselines.

2 Basics and Background

An n-dimensional temporal point process is a random process whose realization consists of a list of discrete events in time and their associated dimension, $\{(t_k, d_k)\}$ with $t_k \in \mathbb{R}^+$ and $d_k \in \{1, \ldots, n\}$. Many different types of data produced in online social networks can be represented as temporal point processes, such as likes and tweets. A temporal point process can be equivalently represented as a counting process, $\mathcal{N}(t) = (\mathcal{N}^1(t), \ldots, \mathcal{N}^n(t))^\top$ associated to n users in the social network. Here, $\mathcal{N}^i(t)$ records the number of events user i performs before time t for $1 \leq i \leq n$. Let the history $\mathcal{H}^i(t)$ be the list of times of events $\{t_1, t_2, \ldots, t_k\}$ of the i-th user up to time t. Then, the number of observed events in a small time window $[t, t + dt)$ of length dt is $d\mathcal{N}^i(t) = \sum_{t_k \in \mathcal{H}^i(t)} \delta(t - t_k)\, dt$, and hence $\mathcal{N}^i(t) = \int_0^t d\mathcal{N}^i(s)$, where $\delta(t)$ is a Dirac delta function. The point process representation of temporal data is fundamentally different from the discrete time representation typically used in social network analysis. It directly models the time interval between events as random variables, avoids the need to pick a time window to aggregate events, and allows temporal events to be modeled in a fine grained fashion. Moreover, it has a remarkably rich theoretical support [6].

An important way to characterize temporal point processes is via the conditional intensity function — a stochastic model for the time of the next event given all the times of previous events. Formally, the conditional intensity function $\lambda^i(t)$ (intensity, for short) of user i is the conditional probability of observing an event in a small window $[t, t + dt)$ given the history $\mathcal{H}(t) = \{\mathcal{H}^1(t), \ldots, \mathcal{H}^n(t)\}$:

$$\lambda^i(t)dt := \mathbb{P}\left\{\text{user } i \text{ performs event in } [t, t + dt) \,|\, \mathcal{H}(t)\right\} = \mathbb{E}[d\mathcal{N}^i(t) \,|\, \mathcal{H}(t)], \tag{1}$$

where one typically assumes that only one event can happen in a small window of size dt. The functional form of the intensity $\lambda^i(t)$ is often designed to capture the phenomena of interests.

The Hawkes process [7] is a class of self and mutually exciting point process models,

$$\lambda^i(t) = \mu^i(t) + \sum_{k: t_k < t} \phi^{id_k}(t, t_k) = \mu^i(t) + \sum_{j=1}^{n} \int_0^t \phi^{ij}(t, s) d\mathcal{N}^j(s), \tag{2}$$

where the intensity is history dependent. $\phi^{ij}(t, s)$ is the impact function capturing the temporal influence of an event by user j at time s to the future events of user j at time $t \geqslant s$. Here, the first term $\mu^i(t)$ is the exogenous event intensity modeling drive outside the network and indecent of the history, and the second term $\sum_{k: t_k < t} \phi^{id_k}(t, t_k)$ is the endogenous event intensity modeling interactions within the network [8]. Defining $\Phi(t, s) = [\phi^{ij}(t, s)]_{i,j=1\ldots n}$, and $\lambda(t) = (\lambda^1(t), \ldots, \lambda^n(t))^\top$, and $\mu(t) = (\mu^1(t), \ldots, \mu^n(t))^\top$ we can compactly rewrite Eq 2 in matrix form:

$$\lambda(t) = \mu(t) + \int_0^t \Phi(t, s) d\mathcal{N}(s). \tag{3}$$

In practice it is standard to employ shift-invariant impact function, *i.e.*, $\Phi(t, s) = \Phi(t - s)$. Then, by using notation of convolution $f(t) * g(t) = \int_0^t f(t - s) g(s) ds$ we have

$$\lambda(t) = \mu(t) + \Phi(t) * d\mathcal{N}(t). \tag{4}$$

3 From Intensity to Average Activity

In this section we will develop a closed form relation between the expected total intensity $\mathbb{E}[\lambda(t)]$ and the intensity $\mu(t)$ of exogenous events. This relation establish the basis of our campaigning framework. First, define the *mean function* as $\mathcal{M}(t) := \mathbb{E}[\mathcal{N}(t)] = \mathbb{E}_{\mathcal{H}(t)}[\mathbb{E}(\mathcal{N}(t)|\mathcal{H}(t))]$. Note that $\mathcal{M}(t)$ is history independent, and it gives the average number of events up to time t for each of the dimension. Similarly, the *rate* function $\eta(t)$ is given by $\eta(t)dt := d\mathcal{M}(t)$. On the other hand,

$$d\mathcal{M}(t) = d\mathbb{E}[\mathcal{N}(t)] = \mathbb{E}_{\mathcal{H}(t)}[\mathbb{E}(d\mathcal{N}(t)|\mathcal{H}(t))] = \mathbb{E}_{\mathcal{H}(t)}[\lambda(t)|\mathcal{H}(t)]dt = \mathbb{E}[\lambda(t)]dt. \tag{5}$$

Therefore $\eta(t) = \mathbb{E}[\lambda(t)]$ which serves as a measure of activity in the network. In what follows we will find an analytical form for the average activity. Proofs are presented in Appendix C.

Lemma 1. *Suppose $\Psi : [0,T] \to \mathbb{R}^{n\times n}$ is a non-increasing matrix function, then for every fixed constant intensity $\mu(t) = c \in \mathbb{R}^n_+$, $\eta_c(t) := \Psi(t)c$ solves the semi-infinite integral equation*

$$\eta(t) = c + \int_0^t \Phi(t-s)\eta(s)ds, \quad \forall t \in [0,T], \tag{6}$$

if and only if $\Psi(t)$ satisfies

$$\Psi(t) = I + \int_0^t \Phi(t-s)\Psi(s)ds, \quad \forall t \in [0,T]. \tag{7}$$

In particular, if $\Phi(t) = Ae^{-\omega t}\mathbf{1}_{\geq 0}(t) = [a_{ij}e^{-\omega t}\mathbf{1}_{\geq 0}(t)]_{ij}$ where $0 \leq \omega \notin$ Spectrum(A), then

$$\Psi(t) = e^{(A-\omega I)t} + \omega(A-\omega I)^{-1}(e^{(A-\omega I)t} - I) \tag{8}$$

for $t \in [0,T]$, where, $\mathbf{1}_{\geq 0}(t)$ is an indicator function for $t \geq 0$.

Let $\mu : [0,T] \to \mathbb{R}^n_+$ be a right-continuous piecewise constant function

$$\mu(t) = \sum_{m=1}^{M} c_m \mathbf{1}_{[\tau_{m-1},\tau_m)}(t), \tag{9}$$

where $0 = \tau_0 < \tau_1 < \cdots < \tau_M = T$ is a finite partition of time interval $[0,T]$ and function $\mathbf{1}_{[\tau_{m-1},\tau_m)}(t)$ indicates $\tau_{m-1} \leq t < \tau_m$. The next theorem shows that if $\Psi(t)$ satisfies (7), then one can calculate $\eta(t)$ for piecewise constant intensity $\mu : [0,T]$ of form (9).

Theorem 2. *Let $\Psi(t)$ satisfy* (7) *and $\mu(t)$ be a right-continuous piecewise constant intensity function of form* (9)*, then the rate function $\eta(t)$ is given by*

$$\eta(t) = \sum_{k=0}^{m} \Psi(t-\tau_k)(c_k - c_{k-1}), \tag{10}$$

for all $t \in (\tau_{m-1}, \tau_m]$ and $m = 1, \ldots, M$, where $c_{-1} := 0$ by convention.

Using the above lemma, for the first time, we derive the average intensity for a general exogenous intensity. Appendix E includes a few experiments to investigate these results empirically.

Theorem 3. *If $\Psi \in C^1([0,T])$ and satisfies* (7)*, and exogenous intensity μ is bounded and piecewise absolutely continuous on $[0,T]$ where $\mu(t+) = \mu(t)$ at all discontinuous points t, then μ is differentiable almost everywhere, and the semi-indefinite integral*

$$\eta(t) = \mu(t) + \int_0^t \Phi(t-s)\eta(s)ds, \quad \forall t \in [0,T], \tag{11}$$

yields a rate function $\eta : [0,T] \to \mathbb{R}^n_+$ given by

$$\eta(t) = \int_0^t \Psi(t-s)d\mu(s). \tag{12}$$

Corollary 4. *Suppose Ψ and μ satisfy the same conditions as in Thm. 3, and define $\psi = \Psi'$, then the rate function is $\eta(t) = (\psi * \mu)(t)$. In particular, if $\Phi(t) = Ae^{-\omega t}\mathbf{1}_{\geq 0}(t) = [a_{ij}e^{-\omega t}\mathbf{1}_{\geq 0}(t)]_{ij}$ then the rate function $\eta(t) = \mu(t) + A\int_0^t e^{(A-wI)(t-s)}\mu(s)ds$.*

4 Multi-stage Closed-loop Control Problem

Given the analytical relation between exogenous intensity and expected overall intensity (rate function), one can solve a single one-stage campaigning problem to find the optimal constant intervention intensity [8]. Alternatively, the time window can be partitioned into multiple stages and one can impose different levels of interventions in these stages. This yields an open-loop optimization of the cost function where one selects all the intervention actions at initial time 0. More effectively, we tackle the campaigning problem in a dynamic and adaptive manner where we can postpone deciding the intervention by observing the process until the next stage begins. This is called the *closed-loop* optimization of the objective function.

In this section, we establish the foundation to formulate the problem as a multi-stage closed-loop optimal control problem. We assume that n users are generating events according to multi-dimensional Hawkes process with exogenous intensity $\mu(t) \in \mathbb{R}^n$ and impact function $\Phi(t,s) \in \mathbb{R}^{n\times n}$.

Event exposure. Event exposure is the quantity of major interests in campaigning. The exposure process is mathematically represented as a counting process, $\mathcal{E}(t) = (\mathcal{E}^1(t), \dots, \mathcal{E}^n(t))^\top$: Here, $\mathcal{E}^i(t)$ records the number of times user i is exposed (she or one of her neighbors performs an activity) to the campaign by time t. Let B be the adjacency matrix of the user network, *i.e.*, $b_{ij} = 1$ if user i follows user j or equivalently user j influences user i. We assume $b_{ii} = 1$ for all i. Then the exposure process is given by $\mathcal{E}(t) = B\,\mathcal{N}(t)$.

Stages and interventions. Let $[0,T]$ be the time horizon and $0 = \tau_0 < \tau_1 < \dots < \tau_{M-1} < \tau_M = T$ be a partition into the M stages. In order to steer the activities of network towards a desired level (criteria given below) at these stages, we impose a constant intervention $u_m \in \mathbb{R}^n$ to the existing exogenous intensity μ during time $[\tau_m, \tau_{m+1})$ for each stage $m = 0, 1, \dots, M-1$. The activity intensity at the m-th stage is $\lambda_m(t) = \mu + u_m + \int_0^t \Phi(t,s)\, d\mathcal{N}(s)$ for $\tau_m \le t < \tau_{m+1}$ where $\mathcal{N}(t)$ tracks the counting process of activities since $t = 0$. Note that the intervention itself exhibits a stochastic nature: adding u_m^i 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 real-world campaigning.

States and state evolution. Note that the Hawkes process is non-Markov and one needs complete knowledge of the history to characterize the entire process. However, the conditional intensity $\lambda(t)$ only depends on the state of process at time t when the standard exponential kernel $\Phi(t,s) = Ae^{-\omega(t-s)}\mathbf{1}_{\ge 0}(t-s)$ is employed. In this case, the activity rate at stage m is

$$\lambda_m(t) = \mu + u_m + \underbrace{\int_0^{\tau_m} Ae^{-\omega(t-s)}\, d\mathcal{N}(s)}_{\text{from previous stages}} + \underbrace{\int_{\tau_m}^{t} Ae^{-\omega(t-s)}\, d\mathcal{N}(s)}_{\text{current stage}} \tag{13}$$

Define $x_m := \lambda_{m-1}(\tau_m) - u_{m-1} - \mu$ (and $x_0 = 0$ by convention) then the intensity due to events of all previous m stages can be written as $\int_0^{\tau_m} Ae^{-\omega(t-s)}\, d\mathcal{N}(s) = x_m e^{-\omega(t-\tau_m)}$. In other words, x_m is sufficient to encode the information of activity in the past m stages that is relevant to future. This is in sharp contrast to the general case where the state space grows with the number of events.

Objective function. For a sequence of controls $u(t) = \sum_{m=0}^{M-1} u_m \mathbf{1}_{[\tau_m,\tau_{m+1})}(t)$, the activity counting process $\mathcal{N}(t)$ is generated by intensity $\lambda(t) = \mu + u(t) + \int_0^t Ae^{-\omega(t-s)}\, d\mathcal{N}(s)$. For each stage m from 0 to $M-1$, x_m encodes the effects from previous m stages as above and u_m is the current control imposed at this stage. Let $\mathcal{E}_m^i(t; x_m, u_m) := B\int_{\tau_m}^t d\mathcal{N}^i(s)$ be the number of times user i is exposed to the campaign by time $t \in [\tau_m, \tau_{m+1})$ in stage m, then the goal is to steer the expected total number of exposure $\bar{\mathcal{E}}_m^i(x_m, u_m) := \mathbb{E}[\mathcal{E}_m^i(\tau_{m+1}; x_m, u_m)]$ to a desired level. In what follows, we introduce several instances of the objective function $g(x_m, u_m)$ in terms of $\{\bar{\mathcal{E}}_m^i(x_m, u_m)\}_{i=1}^n$ in each stage m that characterize different *exposure shaping* tasks. Then the overall control problem is to find $u(t)$ that optimizes the total objective $\sum_{m=0}^{M-1} g_m(x_m, u_m)$.

- *Capped Exposure Maximization (CEM)*: In real networks, there is a cap on the exposure each user can tolerate due to the limited attention of a user. Suppose we know the upper bound β_m^i , on user i's exposure tolerance over which the extra exposure is not counted towards the objective. Then, we can form the following *capped exposure maximization*

$$g_m(x_m, u_m) = \frac{1}{n}\sum_{i=1}^n \min\left\{\bar{\mathcal{E}}_m^i(x_m, u_m), \beta_m^i\right\} \tag{14}$$

Algorithm 1: Closed-loop Multi-stage Dynamic Programming

Input: Intervention constraints: $c_0 \dots c_{M-1}, C_0 \dots C_{M-1}, \alpha_0 \dots \alpha_{M-1}$,
Input: Objective-specific constraints: $\beta_0 \dots \beta_{M-1}$ for CEM and $\gamma_0 \dots \gamma_{M-1}$ for LES
Input: Time: T, Hawkes parameters: A, ω
Output: Optimal intervention $u_0 \dots u_{M-1}$, Optimal cost: $Cost$
Set $x_0 \leftarrow 0$ and $Cost \leftarrow 0$
for $l \leftarrow 0 : M-1$ **do**
  $(v_l \dots v_{M-1})$ = *open_loop*(x_l) (Problems (24), (25), (26) for CEM, MEM, LES respectively)
  Set $u_l \leftarrow v_l$ and drop $v_{l+1} \dots v_{M-1}$
  Update next state $x_{l+1} \leftarrow f_l(x_l, u_l)$ and $Cost = Cost + g_l(x_l, u_l)$

- *Minimum Exposure Maximization (MEM)*: Suppose our goal is instead to maintain the exposure of campaign on each user above a certain minimum level, at each stage or, alternatively to make the user with the minimum exposure as exposed as possible, we can consider the following cost function:

$$g_m(x_m, u_m) = \min_i \bar{\mathcal{E}}^i_m(x_m, u_m) \tag{15}$$

- *Least-squares Exposure Shaping (LES)*: Sometimes we want to achieve a pre-specified target exposure levels, $\gamma_m \in \mathbb{R}^n$, for the users. For example, we may like to divide users into groups and desire a different level of exposure in each group. To this end, we can perform least-squares campaigning task with the following cost function where D encodes potentially additional constraints (e.g., group partitions):

$$g_m(x_m, u_m) = -\frac{1}{n}\|D\bar{\mathcal{E}}_m(x_m, u_m) - \gamma_m\|^2 \tag{16}$$

Policy and actions. By observing the counting process in previous stages (summarized in a sequence of x_m) and taking the future uncertainty into account, the control problem is to design a policy $\pi = \{\pi_m : \mathbb{R}^n \to \mathbb{R}^n : m = 0, \dots, M-1\}$ such that the controls $u_m = \pi_m(x_m)$ can maximize the total objective $\sum_{m=0}^{M-1} g_m(x_m, u_m)$. In addition, we may have constraints on the amount of control. For example, 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 $\mathcal{U}_m := \left\{u_m \in \mathbb{R}^n | c_m^\top u_m \le C_m, 0 \leqslant u_m \leqslant \alpha_m\right\}$. Here, $c_m \in \mathbb{R}^n_+$ contains the price of each person per unit increase of exogenous intensity and $C_m \in \mathbb{R}_+$ is the total budget at stage m. Also, $\alpha_m \in \mathbb{R}^n_+$ is the cap on the amount of activities of the users.

To summarize, the following problem is formulated to find the optimal control policy π:

$$\underset{\pi}{\text{maximize}} \sum_{m=0}^{M-1} g_m(x_m, \pi_m(x_m)), \quad \text{subject to } \pi_m(x_m) \in \mathcal{U}_m, \text{ for } m = 0, \dots, M-1. \tag{17}$$

5 Closed-loop Dynamic Programming Solution

We have formulated the control problem as an optimization in (17). However, when control policy π_m is to be implemented, only x_m is observed and there are still uncertainties in future $\{x_{m+1}, \dots, x_{M-1}\}$. For instance, when π_m is implemented according to x_m starting from time τ_m, the intensity $x_{m+1} := f(x_m, \pi_m(x_m))$ at time τ_{m+1} depends on x_m and the control $\pi_m(x_m)$, but is also random due to the stochasticity of the process during time $[\tau_m, \tau_{m+1})$. Therefore, the design of π needs to take future uncertainties into considerations.

Suppose we have arrived at stage M at time τ_{M-1} with observation x_{M-1}, then the optimal policy π_{M-1} satisfies $g_{M-1}(x_{M-1}, \pi_{M-1}(x_{M-1})) = \max_{u \in \mathcal{U}_{M-1}} g_{M-1}(x_{M-1}, u) =: J_{M-1}(x_{M-1})$. We then repeat this procedure for m from $M-1$ to 0 backward to find the sequence of controls via dynamic programming such that the control $\pi_m(x_m) \in \mathcal{U}_m$ yields optimal objective value

$$J_m(x_m) = \max_{u_m \in \mathcal{U}_m} \mathbb{E}[g_m(x_m, u_m) + J_{m+1}(f(x_m, u_m))] \tag{18}$$

Approximate Dynamic Programming. Solving (18) for finding $J_m(x_m)$ analytically is intractable. Therefore, we will adopt an approximate dynamic programming scheme. In fact approximate control is as essential part of dynamic programming as the optimization is usually intractable due to curse of dimensionality except a few especial cases [3]. Here we adopt a suboptimal control scheme, *certainty equivalent control* (CEC), which applies at each stage the control that would be optimal if the uncertain quantities were fixed at some typical values like the average behavior. It results in an optimal control sequence, the first component of which is used at the current stage, while the remaining components are discarded. The procedure is repeated for the remaining stages. Algorithm 1 summarizes the dynamic programing steps. This algorithm has two parts: (i) certainty equivalence which the random behavior is replaced by its average; and (ii) the open-loop optimization. Let's assume we are at the beginning of stage l of the Alg. 1 with state vector x_l at τ_l.

Certainty equivalence. We use the machinery developed in Sec. 3 to compute the average of exposure at any stage $m = l, l+1, \ldots, M-1$.

$$\bar{\mathcal{E}}_m(x_m, u_m) = B\mathbb{E}[\mathcal{N}(\tau_{m+1}) - \mathcal{N}(\tau_m)] = B\mathbb{E}\left[\int_{\tau_m}^{\tau_{m+1}} d\mathcal{N}(s)\right] = B\int_{\tau_m}^{\tau_{m+1}} \eta_m(s)\, ds \quad (19)$$

where $\eta_m(t) = \mathbb{E}[\lambda_m(t)]$ and $\lambda_m(t) = \mu + u_m + x_l e^{-\omega(t-\tau_l)} + \int_{\tau_l}^{t} Ae^{-\omega(t-s)} d\mathcal{N}(s)$ for $t \in [\tau_m, \tau_{m+1})$. Now, we use the superposition property of point processes [4] to decompose the process as $\mathcal{N}(t) = \mathcal{N}^c(t) + \mathcal{N}^v(t)$ corresponding to $\lambda_m(t) = \lambda_m^c(t) + \lambda_m^v(t)$ where the first $\lambda_m^c(t) = \mu + u_m + \int_{\tau_l}^{t} Ae^{-\omega(t-s)} d\mathcal{N}^c(s)$ consists of events caused by exogenous intensity at current stage m and the second $\lambda_m^v(t) = x_l e^{-\omega(t-\tau_l)} + \int_{\tau_l}^{t} Ae^{-\omega(t-s)} d\mathcal{N}^v(s)$ is due to activities in previous stages. According to Thm. 2 we have

$$\eta_m^c(t) := \mathbb{E}[\lambda_m^c(t)] = \Psi(t-\tau_l)\mu + \Psi(t-\tau_l)u_l + \sum_{k=l+1}^{m-1} \Psi(t-\tau_k)(u_k - u_{k-1}), \quad (20)$$

and according to Thm. 3 we have

$$\eta_m^v(t) := \mathbb{E}[\lambda_m^v(t)] = \int_{\tau_l}^{t} \Psi(t-s)\, d(x_l e^{-\omega(s-\tau_l)} \mathbf{1}_{[\tau_l,\infty)}(s)). \quad (21)$$

From now on, for simplicity, we assume stages are based on equal partition of $[0, T]$ to M segments where each has length Δ_M. Combining Eq. (19) and $\eta_m(t) = \eta_m^c(t) + \eta_m^v(t)$ yields:

$$\begin{aligned} \bar{\mathcal{E}}_m(x_m, u_m) =& \Gamma((m-l+1)\Delta_M)u_l + \Gamma((m-l)\Delta_M)(u_{l+1} - u_l) + \ldots \\ &+ \Gamma(\Delta_M)(u_m - u_{m-1}) + \Gamma((m-l+1)\Delta_M)\mu + \Upsilon((m-l+1)\Delta_M)x_l \end{aligned} \quad (22)$$

where $\Gamma(t)$ and $\Upsilon(t)$ are matrices independent of u_m's and are defined in Appendix D. Note the linear relation between average exposure $\bar{\mathcal{E}}_m(x_m, u_m)$ and intervention values $u_l, \ldots, u_{m-1}$.

Open-loop optimization. Having found the average exposure at stages $m = l, \ldots, M-1$ we formulate an open-loop optimization to find optimal $u_l, u_{l+1}, \ldots, u_{M-1}$. Defining $\hat{u}_l = (u_l; \ldots; u_{M-1})$ and $\hat{\mathcal{E}}_l = (\bar{\mathcal{E}}_l(x_l, u_l); \ldots; \bar{\mathcal{E}}_{M-1}(x_{M-1}, u_{M-1}))$ we can write

$$X_l \hat{u}_l + Y_l \mu + W_l x_l = \hat{\mathcal{E}}_l \quad \text{where} \quad Z_l \hat{u}_l \leq z_l \quad (23)$$

and X_l, Y_l, W_l, Z_l, and z_l are independent of $\hat{u}_l$, μ, and x_l as defined in Appendix D.

Defining the expanded form of constraint variables as $\hat{c}_l = (c_l; \ldots; c_{M-1})$, $\hat{C}_l = (C_l; \ldots; C_{M-1})$, and $\hat{\alpha}_l = (\alpha_l; \ldots; \alpha_{M-1})$ we provide the optimization from of the above exposure shaping tasks.

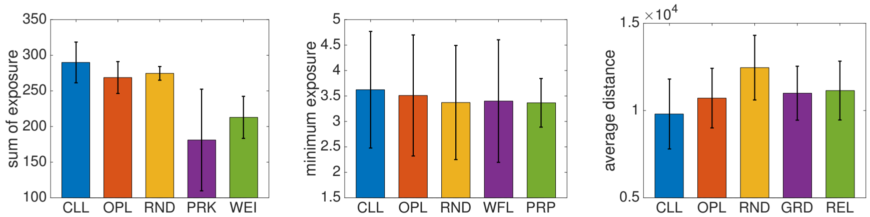
For CEM consider $\hat{\beta}_l = (\beta_l; \ldots, \beta_{M-1})$. Then the problem

$$\text{maximize}_{\hat{h}, \hat{u}_l} \tfrac{1}{n}\mathbf{1}^\top \hat{h} \quad \text{subject to } X_l \hat{u}_l + Y_l \mu + W_l x_l \geq \hat{h},\ \ \hat{\beta}_l \geq \hat{h},\ \ Z_l \hat{u}_l \leq z_l, \quad (24)$$

solves CEM where h is an auxiliary vector of size $n(M-l)$.

For MEM consider the auxiliary h as a vector of size $M-l$ and $\hat{h}$ a vector of size $n(M-1)$. $\hat{h} = (h(1); \ldots; h(1); h(2); \ldots, h(2); \ldots, h(M-l); \ldots; h(M-l))$ where each $h(k)$ is repeated n times. Then MEM is equivalent to

$$\text{maximize}_{\hat{h}, \hat{u}_l} \mathbf{1}^\top \hat{h} \quad \text{subject to } X_l \hat{u}_l + Y_l \mu + W_l x_l \geq \hat{h},\ \ \hat{\beta}_l \geq \hat{h},\ \ Z_l \hat{u}_l \leq z_l \quad (25)$$



a) Capped maximization b) Minimum maximization c) Least-squares shaping

Figure 1: The objective on simulated events and synthetic network; $n = 300$, $M = 6$, $T = 40$

For LES let $\hat{\gamma}_l = (\gamma_l; \ldots; \gamma_{M-1})$ and $\hat{D}_l = diag(D, \ldots, D)$, then

$$\text{minimize}_{\hat{u}_l} \frac{1}{n} \|\hat{D}_l(X_l \hat{u}_l + Y_l \mu + W_l x_l) - \hat{\gamma}_l\|^2 \quad \text{subject to } Z_l \hat{u}_l \leq z_l \tag{26}$$

All the three tasks involve convex (and linear) objective function with linear constraints which impose a convex feasible set. Therefore, one can use the rich and well-developed literature on convex optimization and linear programming to find the optimum intervention.

6 Experiments

We evaluate our campaigning framework using both simulated and real world data and show that our approach significantly outperforms several baselines[1].

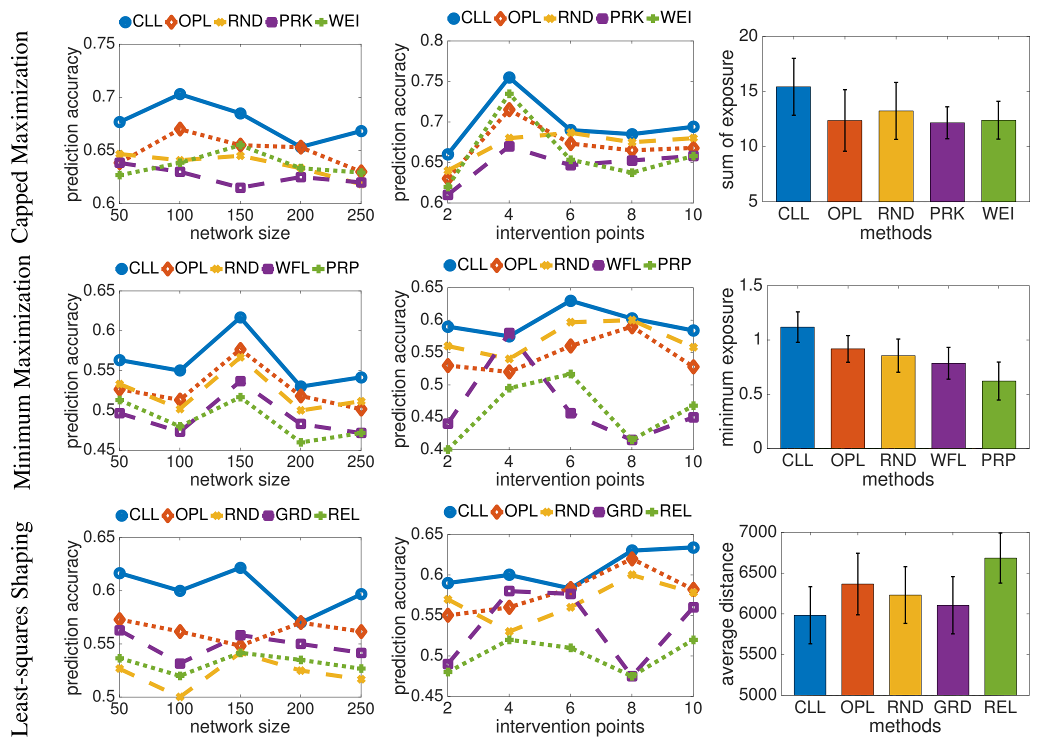
Campaigning results on synthetic networks. In this section, we experiment with a synthetic network of 300 nodes. Details of the experimental setup and parameter setting are found in appendix F. We focus on three tasks: capped exposure maximization, minimax exposure shaping, and least square exposure shaping. To compare the methods we simulate the network with the prescribed intervention intensity and compute the objective function based on the events happened during the simulation. The mean and standard deviation of the objective function out of 10 runs are reported.

Fig. 1 summarizes the performance of the proposed algorithm (CLL) and 4 other baselines on different campaigning tasks. For **CEM**, our approach consistently outperforms the others by at least 10. This means it exposes each user to the campaign at least 10 times more than the rest consuming the same budget and within the same constraints. The extra 20 units of exposures of over OPL or value of information shows how much we gain by incorporating a dynamic closed-loop solution as opposed to open-loop one-time optimization over all stages. For **MEM**, the proposed method outperforms the others by a smaller margin, however, the 0.1 exposure difference with the second best method is not trifling. This is expected as lifting the minimum exposure is a difficult task [8]. For **LES**, results demonstrate the superiority of CLL by a large margin. The 10^3 difference with the second best algorithm aggregated over 6 stages roughly is translated to $\sqrt{10^3/6} \sim 13$ difference in the number of exposures per user. Given the heterogeneity of the network activity and target shape, this is a significant improvement over the baselines. Appendix F includes further results on varying number of nodes, number of stages, and duration of each stage.

Campaigning results on real world networks. We also evaluate the proposed framework on real world data. To this end, we utilize the MemeTracker dataset [9] which contains the information flows captured by hyperlinks between different sites with timestamps during 9 months. This data has been previously used to validate Hawkes process models of social activity [5, 10]. For the real data, we utilize two evaluation procedures. First, similar to the synthetic case, we simulate the network, but now on a network based on the learned parameters from real data. However, the more interesting evaluation scheme would entail carrying out real intervention in a social media platform. Since this is very challenging to do, instead, in this evaluation scheme we used held-out data to mimic such procedure. Second, we form 10 pairs of clusters/cascades by selecting any 2 combinations of 5 largest clusters in the Memetracker data. Each is a cascade of events around a common subject. For any of these 10 pairs, the methods are faced to the question of predicting which cascade will reach the objective function better. They should be able to answer this by measuring how similar their prescription is to the real exogenous intensity. The key point here is that the real events happened are used to evaluate the objective function of the methods. Then the results are reported on average prediction accuracy on all stages over 10 runs of random constraint and parameter initialization on 10 pairs of cascades. The details of the experimental setup is further explained in Appendix F.

Fig. 2, left column illustrates the performance with respect to increasing the number of users in the network. The performance drops slightly with the network size. This means that prediction becomes

[1] codes are available at `http://www.cc.gatech.edu/~mfarajta/`



Performance vs. # users   Performance vs. # points   Objective function

Figure 2: real world dataset results; $n = 300$, $M = 6$, $T = 40$

more difficult as more random variables are involved. The middle panel shows the performance with respect to increasing the number of intervention points. Here, a slight increase in the performance is apparent. As the number of intervention points increases the algorithm has more control over the outcome and can reach the objective function better.

Fig. 2 top row summarizes the results of **CEM**. The left panel demonstrates the predictive performance of the algorithms. CLL consistently outperforms the rest. With 65-70 % of accuracy in predicting the optimal cascade. The right panel shows the objective function simulated 10 times with the learned parameters for network of $n = 300$ users on 6 intervention points. The extra 2.5 extra exposure per user compared to the second best method with the same budget and constraint would be a significant advertising achievement. Among the competitors OPL and RND seem to perform good. If there where no cap over the resultant exposure, all methods would perform comparably because of the linearity of sum of exposure. However, the successful method is the one who manage to maximize exposure considering the cap. Failure of PRK and WEI indicates that structural properties are not enough to capture the influence. Compared to these two, RND performs better in average, however exhibits a larger variance as expected.

Fig. 2 middle row summarizes the results for **MEM** and shows CLL outperforms others consistently. CLL still is the best algorithm and OPL and RND are the significant baselines. Failure of WFL and PRP shows the network structure plays a significant role in the activity and exposure processes.

The bottom row in Fig. 2 demonstrates the results of **LES**. CLL is still the best method. OPL is still strong but RND is not performing well. The objective function is summation of the square of the gap between target and current exposure. This explains why GRD is showing a comparable success, since, it starts with the highest gap in the exposure and greedily allocates the budget.

Conclusion. In this paper, we introduced the optimal multistage campaigning problem, which is a generalization of the activity shaping and influence maximization problems, and it allows for more elaborate goal functions. Our model of social activity is based on multivariate Hawkes process, and for the first time, we manage to derive a linear connection between a time-varying exogenous intensity and the overall network exposure of the campaign.

Acknowledgement. The work is supported in part by NSF/NIH BIGDATA R01 GM108341, NSF IIS-1639792, NSF DMS-1620345, and NSF DMS-1620342.

References

[1] D M West. *Air Wars: Television Advertising and Social Media in Election Campaigns, 1952-2012: Television Advertising and Social Media in Election Campaigns, 1952-2012*. Sage, 2013.

[2] M Vergeer, L Hermans, and S Sams. Online social networks and micro-blogging in political campaigning the exploration of a new campaign tool and a new campaign style. *Party Politics*, 2013.

[3] Dimitri P Bertsekas. *Dynamic programming and optimal control*, volume 1.

[4] D J Daley and D Vere-Jones. *An introduction to the theory of point processes*. Springer Science & Business Media, 2007.

[5] K Zhou, H Zha, and L Song. Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In *AISTATS*, 2013.

[6] O Aalen, O Borgan, and H Gjessing. *Survival and event history analysis: a process point of view*. Springer, 2008.

[7] A G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 1971.

[8] M Farajtabar, N Du, M Gomez-Rodriguez, I Valera, L Song, and H Zha. Shaping social activity by incentivizing users. NIPS, 2014.

[9] J Leskovec, L Backstrom, and J Kleinberg. Meme-tracking and the dynamics of the news cycle. SIGKDD, 2009.

[10] SH Yang and H Zha. Mixture of mutually exciting processes for viral diffusion. ICML, 2013.

[11] D Kempe, J Kleinberg, and E Tardos. Maximizing the spread of influence through a social network. SIGKDD, 2003.

[12] FB Hanson. *Applied stochastic processes and control for Jump-diffusions: modeling, analysis, and computation*, volume 13. Siam, 2007.

[13] A De, I Valera, N Ganguly, S Bhattacharya, and M Gomez Rodriguez. Modeling opinion dynamics in diffusion networks. *arXiv:1506.05474*, 2015.

[14] Y Wang, E Theodorou, A Verma, and L Song. Steering opinion dynamics in information diffusion networks. *arXiv:1603.09021*, 2016.

[15] D Bloembergen, B Ranjbar Sahraei, H Bou-Ammar, K Tuyls, and G Weiss. Influencing social networks: An optimal control study. In *ECAI*, 2014.

[16] K Kandhway and J Kuri. Campaigning in heterogeneous social networks: Optimal control of si information epidemics. 2015.

[17] Pin-Yu Chen, Shin-Ming Cheng, and Kwang-Cheng Chen. Optimal control of epidemic information dissemination over networks. *Cybernetics, IEEE Transactions on*, 2014.

[18] W Lian, R Henao, V Rao, J Lucas, and L Carin. A multitask point process predictive model. ICML, 2015.

[19] AP Parikh, A Gunawardana, and C Meek. Conjoint modeling of temporal dependencies in event streams. UAI, 2012.

[20] PO Perry and PJ Wolfe. Point process modeling for directed interaction networks. *Journal of the Royal Statistical Society*, 2013.

[21] SW Linderman and RP Adams. Discovering latent network structure in point process data. ICML, 2014.

[22] C Blundell, J Beck, and KA Heller. Modelling reciprocating relationships with hawkes processes. NIPS, 2012.

[23] T Iwata, A Shah, and Z Ghahramani. Discovering latent influence in online social activities via shared cascade poisson processes. SIGKDD, 2013.

[24] O Hijab. *Introduction to calculus and classical analysis*. Springer, 2007.

[25] GB Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 2013.

[26] R Bracewell. The fourier transform and iis applications. *New York*, 5, 1965.

[27] AH Al-Mohy and MJ Higham. Computing the action of the matrix exponential, with an application to exponential integrators. *SIAM journal on scientific computing*, 2011.