Demarcating Endogenous and Exogenous Opinion Diffusion Process on Social Networks

Abir De*
IIT Kharagpur
abir.de@cse.iitkgp.ernet.in

Sourangshu Bhattacharya
IIT Kharagpur
sourangshu@cse.iitkgp.ernet.in

Niloy Ganguly
IIT Kharagpur
niloy@cse.iitkgp.ernet.in

ABSTRACT

The networked opinion diffusion in online social networks (OSN) is governed by the two genres of opinions–*endogenous* opinions that are driven by the influence of social contacts between users, and *exogenous* opinions which are formed by external effects like news, feeds etc. Such duplex opinion dynamics is led by users belonging to two categories– *organic* users who generally post endogenous opinions and *extrinsic* users who are susceptible to externalities, and mostly post the exogenous messages. Precise demarcation of endogenous and exogenous messages offers an important cue to opinion modeling, thereby enhancing its predictive performance. On the other hand, accurate user selection aids to detect extrinsic users, which in turn helps in opinion shaping. In this paper, we design CherryPick, a novel learning machinery that classifies the opinions and users by solving a joint inference task in message and user set, from a temporal stream of sentiment messages. Furthermore, we validate the efficacy of our proposal from both modeling and shaping perspectives. Moreover, for the latter, we formulate the opinion shaping problem in a novel framework of stochastic optimal control, in which the selected extrinsic users optimally post exogenous messages so as to guide the opinions of others in a desired way. On five datasets crawled from Twitter, CherryPick offers a significant accuracy boost in terms of opinion forecasting, against several competitors. Furthermore, it can precisely determine the quality of a set of control users, which together with the proposed online shaping strategy, consistently steers the opinion dynamics more effectively than several state-of-the-art baselines.

ACM Reference Format:
Abir De, Sourangshu Bhattacharya, and Niloy Ganguly. 2018. Demarcating Endogenous and Exogenous Opinion Diffusion Process on Social Networks. In *WWW 2018: The 2018 Web Conference, April 23–27, 2018, Lyon, France.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3178876.3186121

1 INTRODUCTION

Research on understanding opinion dynamics, from both modeling and control perspectives, abounds in literature, predominantly following two approaches [1–15]. While the first approach that is grounded on the concepts of statistical physics, is barely data-driven and therefore shows poor predictive performance [1–3, 7–15], the second class of models aims to overcome such limitations, by learning a tractable linear model from transient opinion dynamics [4–6]. Barring the individual limitations of these existing approaches, they all have looked into the opinion dynamics phenomenon through the tinted glass of a naive assumption – that of absence or lack of external effects, despite empirical evidences advocating the presence of such signals [16–20]. As a result, the existing models "as it is" only perform modestly in predicting the opinion dynamics.

Since a social network is an open system encouraging both inward and outward flow of information, a continuous flux of external information is funneled to its users, via a gamut of sources like news, feeds, etc. As a result, a networked opinion formation process that involves extensive interactive discussions between connected users, is also propelled by such external sources recommended to those users. Therefore, at the very outset, we observe two families of opinions – *endogenous* opinions which evolve due to the influence from neighbors, and *exogenous* opinions that are driven mostly by the externalities. Such dual dynamics further brackets the users into two categories: *organic* users who predominantly express endogenous opinions, and *extrinsic* users who largely post exogenous contents– together, they organically guide the coupled opinion diffusion process in a social network. In most practical situations, neither true labels of the posts (endogenous or exogenous), nor the users (organic or extrinsic) are available. Therefore, accurate unsupervised labeling of the users and their posts offers a sophisticated trait in opinion modeling– thereby boosting the predictive performance for a broad spectrum of applications like pole-prediction, brand sentiment estimation, etc.

Besides prediction, accurate user classification has an immense potential impact on opinion shaping. An effective user categorization technique helps us spot extrinsic users, *i.e.* people who are susceptible to external posts. Such users can be easily actuated in an *opinion shaping* task, in which feeds or news are posted on their walls, in order to steer the opinions of others to a given state. The task of opinion shaping has been taken up recently by [7–14, 21], however none of the existing opinion shaping approaches aims to identify the extrinsic users, which renders the control strategies practically ineffective. Moreover, the approach in [21] assigns the control signals to each and every node, which in essence means that each user is a control user who governs the opinions of others; consequently, their proposal falls wayside of any practical importance. A couple of recent works [22, 23] adopt a similar direction, which, however, focus on entirely different applications, e.g. smart broadcasting and activity maximization.

In this paper, our goal is to demarcate endogenous and exogenous messages, classify organic and extrinsic users, and finally demonstrate the utility of our proposal both from opinion modeling and *especially opinion shaping* viewpoints, where for the latter, we devise an efficient control mechanism, in order to curate the overall opinion dynamics in a favorable manner.

*Now affiliated to MPI for Software Systems, Germany. Email: ade@mpi-sws.org

WWW 2018, April 23–27, 2018, Lyon, France

ACM ISBN 978-1-4503-5639-8/18/04.
https://doi.org/10.1145/3178876.3186121

Proposed approach: We initiate our investigation by positing the dynamics of organic opinion in the presence of exogenous actions, using our previously proposed model SLANT [6]. It allows users' latent *endogenous* opinions to be modulated over time, by *both endogenous and exogenous* opinions of their neighbors, expressed as sentiment messages (Section 3).

Subsequently, we propose CherryPick, a principled learning mechanism that optimally demarcates the endogenous and exogenous opinions, and classifies organic and extrinsic users. In order to categorize messages as well as the corresponding users, we aim to select the set of events that comply with the organic dynamics with a high confidence, *i.e.* a low variance of influence estimation. To this end, we devise this problem as a joint inference task of both message and user category. We find that this proposed inference problem can be formulated as an instance of cardinality constrained multidimensional submodular maximization problem. To solve this optimization problem, we design a novel greedy approach which too, like an ordinary greedy submodular maximization algorithm, enjoys a $(1-1/e)$ approximation bound (Section 4).

In order to show the efficacy of our user selection approach, we propose an opinion shaping task cast as a novel stochastic optimal control problem. In a marked departure from the prior works, we tackle the shaping problem by decoupling the intensities of the selected extrinsic users, into exogenous ($\boldsymbol{\eta}(t)$) and endogenous parts ($\lambda(t)$), where the exogenous rate is associated with a cost to limit the number of control messages. We find that the optimal value of this multidimensional control signal linearly depends on the current opinion, thus giving a simple, yet scalable closed loop solution to the shaping problem (Section 5).

Finally, we perform experiments on a set of five diverse datasets crawled from Twitter and show that CherryPick by classifying endogenous and exogenous messages, helps in achieving a substantial performance boost in forecasting opinions. Furthermore, we observe that the selected extrinsic users, along with the proposed shaping strategy, consistently steer the opinion dynamics of others more effectively than several baselines (Section 6).

Contributions: Summarizing, our main contributions in this paper are twofold:

1. **An unsupervised demarcation approach:** Our proposal offers CherryPick, a novel *unsupervised* learning algorithm that jointly classifies a stream of unlabeled messages, and their users, in the scenario of opinion dynamics. In principle, CherryPick is a greedy algorithm that maximizes a novel function f, an inverse measure of parameter variance. We find that f enjoys a joint submodular property in both user and message-set, affording provable approximation guarantees from the proposed algorithm. Despite complex inter-dependencies between the message streams, the presence of such an important function, we believe, is a surprising and key observation.

2. **Opinion shaping by actuating extrinsic users:** To establish the utility of extrinsic user identification, we develop a novel stochastic opinion control framework that computes the optimal control message intensities, with which the extrinsic users should post, so as to guide the opinion dynamics in a desired way. In a marked departure from prior works, our proposal offers a closed loop feedback control policy that computes the required message intensities online.

2 RELATED WORK

Opinion modeling and their applications have been widely studied in different guises in many years. In this section, we review some of them, from three major perspectives– (i) opinion dynamics modeling, (ii) opinion sensing, and (iii) opinion shaping.

Opinion dynamics modeling. Modeling the evolution process of opinion flow over networks, mostly follows two approaches, based on (a) statistical physics and (b) data-driven techniques. The first type of models, e.g. Voter, Flocking, DeGroot, etc. is traditionally designed to capture various regulatory real-life phenomena e.g. consensus, polarization, clustering, coexistence etc. [1–3, 24–31]. Voter model [1] is a discrete opinion model, where opinions are represented as nominal values, and copied from influencing neighbors in every step. This underlying principle is still a major workhorse for many discrete opinion models [3, 24–29, 31]. In contrast to these models, Flocking and DeGroot are continuous opinion models. In Flocking model and its variations [30], a node i having opinion x_i first selects the set of neighbors j with $|x_i - x_j| \leq \epsilon$, and then updates its own opinion by averaging these opinions. DeGroot model [2], on the other hand, allows a user to update her opinion with the average opinions of *all* her neighbors. In this model, the underlying influence matrix is row stochastic, enforcing consensus for a strongly connected graph. The second class of models, e.g. Biased Voter, AsLM, SLANT, etc. aims to learn a tractable linear model from a temporal message stream reflecting transient opinion dynamics [4–6]. While a Biased Voter model [4] unifies various aspects of DeGroot and Flocking models, AsLM [5] generalizes the DeGroot model by relaxing the structure of influence matrix. In contrast to these models that ignore the temporal effects of messages (post-rate), SLANT [6] blends the opinion dynamics along with the message dynamics, using a stochastic generative model. However, all these approaches skirt the effect of externalities, which severely constrains their forecasting prowess.

Opinion sensing: Sensing opinions, or mining sentiments from textual data traditionally relies on sophisticated NLP based machineries. See [32, 33] for details. Both these monographs provide a comprehensive survey. In general, LIWC [34] is widely considered as benchmark tool to compute sentiments from rich textual data. On the other hand, Hannak *et al.* developed a simple yet effective method for sentiment mining from short informal text like tweets [35], also used by [5, 6]. Recently, a class of works [36–38] designs simple supervised strategies to sense opinion spams, and some of them [37, 38] also advocates the role of temporal signals on opinion spamming. Note that, exogenous opinions are fundamentally different from opinion spams. In contrast to a spam which is unsolicited and irrelevant to the discussion, an exogenous post is often relevant, yet just an informed reflection of some external news or feeds. Also, since spamminess of a message is its intrinsic property, it does not depend on the messages before it. However, an exogenous post when retweeted, can become endogenous (see Table 3). Furthermore, the opinion spam detection techniques rest on the principle of supervised classification that in turn requires labelled messages. However in the context of networked opinion dynamics, the messages (tweets) come unlabeled, which renders the spam detection techniques practically inapplicable for such scenarios.

Opinion control: Opinion shaping has been studied mostly by the control theorists [7–14]. These works emphasize consensus control, and therefore have limited applicability in most practical scenarios. Furthermore, most of them assume control opinions as continuous signals, whereas in practice, the expressed opinions are discrete events observed only through the messages or posts. Only very recently [21] attempts to overcome these limitations by modeling control signals as discrete epochs, which, however, offers an approximate and computationally inefficient solution.

3 MODEL FORMULATION

In this section, we first revisit the model of opinion dynamics in the absence of exogenous actions [6], and then describe the same in the presence of exogenous actions.

3.1 Problem setup

We use two sources of data as input: a directed social network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of users with the connections between them (e.g. friends, following, etc), and an aggregated history $\mathcal{U}(T)$ of the messages posted by these users during a given time-window $[0, T)$. In this paper, we summarize each message-event $e_i \in \mathcal{U}(T)$ using only three components, the user u_i who has posted the message, the opinion or sentiment value ζ_i associated with the message, and the timestamp t_i of the post. Therefore, $\mathcal{U}(T) := \{e_i = (u_i, \zeta_i, t_i) | t_i < T\}$. We also use $\mathcal{U}(t)$ to denote the set of messages collected until $t < T$.

In a spirit similar to the one proposed in [6], we assume that the history of events until time t influences the arrival process of events after time t. However, in a direct contrast to [6] which skirts the potential influence from externalities, we posit that the message events belong to two categories– *endogenous* and *exogenous*. At the very outset, the arrivals of endogenous events are driven by the previous events in the network, while exogenous events are the rest, originating from external influence, outside the given social network. Note that the distinction between endogenous and exogenous events is not directly observable from the data, but needs to be inferred from the characteristics of the event sequence. To this end, we denote $\mathcal{H}(t)$ and $C(t)$ as the sets of endogenous and exogenous events respectively, observed until time t, with $\mathcal{U}(t) = \mathcal{H}(t) \cup C(t)$. At a user level, we denote $\mathcal{H}_u(t) = \{(u_i, m_i, t_i) | u_i = u \text{ and } t_i < t\}$ as the collection of all endogenous messages with sentiment m_i's, posted by user u until time t. Therefore, $\cup_{u \in \mathcal{V}} \mathcal{H}_u(t) = \mathcal{H}(t)$.

To model the endogenous message dynamics, we represent the message times by a set of counting processes denoted as a vector $\boldsymbol{N}(t)$, in which the u-th entry, $N_u(t) \in \{0\} \cup \mathbb{Z}^+$, counts the number of messages user u posted until time t. Then, we characterize the message rates with the conditional intensity function

$$\mathbb{E}[d\boldsymbol{N}(t) \,|\, \mathcal{U}(t)] = \boldsymbol{\lambda}^*(t)\, dt, \tag{1}$$

where $d\boldsymbol{N}(t) := (\, dN_u(t)\,)_{u \in \mathcal{V}}$ counts the endogenous messages per user in the interval $[t, t + dt)$ and $\boldsymbol{\lambda}^*(t) := (\, \lambda_u^*(t)\,)_{u \in \mathcal{V}}$ denotes the user intensities that depend on the history $\mathcal{U}(t)$. We denote the set of user that u *follows* by $\mathcal{N}(u)$.

Opinion dynamics in absence of exogenous actions [6]: For clarity, we briefly discuss the proposal by De *et al.* [6], that ignores the effect of exogenous messages. The user intensities $\lambda_u^*(t)$ are generally modeled using multivariate Hawkes Process [6, 39–42]. In absence of exogenous actions, *i.e.*, when $\mathcal{U}(t) = \mathcal{H}(t)$, we have:

$$\lambda_u^*(t) = \mu_u + \sum_{v \in \mathcal{N}(u)} b_{vu} \sum_{e_i \in \mathcal{H}_v(t)} \kappa(t - t_i). \tag{2}$$

Here, the first term, $\mu_u \geqslant 0$, captures the posts by user u on her own initiative, and the second term, with $b_{vu} \geqslant 0$, reflects the influence of previous posts on her intensity (self-excitation). The users' latent opinions are represented as a history-dependent, multidimensional stochastic process $\boldsymbol{x}^*(t)$:

$$x_u^*(t) = \alpha_u + \sum_{v \in \mathcal{N}(u)} a_{vu} \sum_{e_i \in \mathcal{H}_v(t)} m_i g(t - t_i) \tag{3}$$

where the first term, $\alpha_u \in \mathbb{R}$, models the original opinion a user u and the second term, with $a_{vu} \in \mathbb{R}$, models updates in user u's opinion due to the influence from previous messages of her neighbors. The influnce values, in practice, may depend on various network properties [43–45]. Here, $\kappa(t) = e^{-\nu t}$ and $g(t) = e^{-\omega t}$ (where $\nu,\ \omega \geqslant 0$) denote exponential triggering kernels, which models the decay of influence over time. Finally, when a user u posts a message at time t, the *message sentiment m reflects the expressed opinion* which is sampled from a distribution $p(m|x_u^*(t))$. Here, the sentiment distribution $p|x_u^*(t)$ is assumed to be normal, *i.e.* $p(m|x_u(t)) = \mathcal{N}(x_u(t), \sigma_u)$.

3.2 Opinion dynamics with exogenous events

As introduced before, $C(t)$ is the collection of exogenous messages posted until time t. Similar to $\mathcal{H}_u(t)$, we also specify $C_u(t) = \{(u_i, w_i, t_i) | t_i < t, u_i = u\}$ as the set of exogenous messages posted by user u until time t. To make a clear distinction, we use m_i and w_i for endogenous and exogenous message sentiments (expressed opinions) respectively. In order to represent the arrival times of the exogenous message set $C(t)$, we introduce an additional counting process $\boldsymbol{M}(t)$ that regulates the rate of publication of the corresponding opinions w_i. Note that, we do not aim to model the dynamics of exogenous events, since their source is not known to us. However, we assume that every exogenous post influences the subsequent endogenous events in the same manner as the previous endogenous events. This is because a recipient user cannot distinguish between exogenous or endogenous posts made by her neighbors. Now, we present the dynamics of latent opinion $x_u^*(t)$ of user u, in the presence of exogenous messages in the following.

$$x_u^*(t) = \alpha_u + \sum_{v \in \mathcal{N}(u)} a_{vu} \Big(\sum_{e_i \in \mathcal{H}_v(t)} m_i g(t - t_i) + \sum_{e_i \in C_v(t)} w_i g(t - t_i) \Big) \tag{4}$$

where, the last term captures signals from exogenous posts. Similarly, the endogenous message rate $\lambda_u^*(t)$ of a user u evolves as,

$$\lambda_u^*(t) = \mu_u + \sum_{v \in \mathcal{N}(u)} b_{vu} \Big(\sum_{e_i \in \mathcal{H}_v(t)} \kappa(t - t_i) + \sum_{e_i \in C_v(t)} \kappa(t - t_i) \Big). \tag{5}$$

Note that same parameters, a_{vu} and b_{vu}, are used to model the effect of endogenous and exogenous processes, on both opinion and message dynamics. The above equation can be equivalently written as:

$$\boldsymbol{x}^*(t) = \boldsymbol{\alpha} + \int_0^t \boldsymbol{A} g(t - s)[\boldsymbol{m}(s) \odot d\boldsymbol{N}(s) + \boldsymbol{w}(s) \odot d\boldsymbol{M}(s)] \tag{6}$$

$$\boldsymbol{\lambda}^*(t) = \boldsymbol{\mu} + \int_0^t \boldsymbol{B} \kappa(t - s)[d\boldsymbol{N}(s) + d\boldsymbol{M}(s)]. \tag{7}$$

Here $\boldsymbol{A} = (a_{vu}) \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$, $\boldsymbol{B} = (b_{vu}) \in \mathbb{R}_+^{|\mathcal{V}|\times|\mathcal{V}|}$, $\boldsymbol{x}^*(t) = (x_u^*(t))_{u\in\mathcal{V}}$. Similarly we define $\boldsymbol{\lambda}^*(t)$, $\boldsymbol{m}(s)$, $\boldsymbol{w}(s)$. Furthermore, the exogenous intensity is given by: $\mathbb{E}[d\boldsymbol{M}(t)|\mathcal{U}(t)] = \boldsymbol{\eta}(t)$. We do not aim to model $\boldsymbol{\eta}(t)$. However, we do utilize it during opinion shaping in section 5.

By defining, $\boldsymbol{P}(t) := \boldsymbol{N}(t) + \boldsymbol{M}(t)$, as the counting process associated with $\mathcal{U}(t)$, we further simplify Eqs. (6) and (7) as,

$$\boldsymbol{x}^*(t) = \boldsymbol{\alpha} + \boldsymbol{A}\int_0^t g(t-s)[\boldsymbol{\zeta}(s)\odot d\boldsymbol{P}(s)] \tag{8}$$

$$\boldsymbol{\lambda}^*(t) = \boldsymbol{\mu} + \int_0^t \kappa(t-s)\boldsymbol{B}\boldsymbol{P}(s). \tag{9}$$

SDE based representation: Given the triggering kernels to be exponential, the resulting opinion and event dynamics are Markovian, and therefore can be represented as jump stochastic differential equations. This representation will be subsequently used in Section 5 for opinion shaping, where the exogenous sentiments $\boldsymbol{w}(t)$ and posts represented by the counting process $\boldsymbol{M}(t)$, will act as the control signals to regulate the endogenous opinions $\boldsymbol{x}^*(t)$.

Proposition 1. *Given the triggering kernel $g(t) = e^{-\omega t}$ and $\kappa(t) = e^{-\nu t}$, the tuple $(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t))$ following Eqs. (6)-(7), is a Markov process, whose dynamics are defined by the following marked jumped stochastic differential equations (SDE):*

$$d\boldsymbol{x}^*(t) = \omega(\boldsymbol{\alpha} - \boldsymbol{x}^*(t))dt + \boldsymbol{A}[\boldsymbol{m}(t)\odot d\boldsymbol{N}(t) + \boldsymbol{w}(t)\odot d\boldsymbol{M}(t)]$$
$$d\boldsymbol{\lambda}^*(t) = \nu(\boldsymbol{\mu} - \boldsymbol{\lambda}^*(t))dt + \boldsymbol{B}\,d\boldsymbol{N}(t) + \boldsymbol{B}d\boldsymbol{M}(t). \tag{10}$$

The proposition can be easily proved by differentiating Eqs. (6) and (7) respectively. A formal proof is given in [46].

4 DEMARCATION OF MESSAGES AND USERS

In this section, we propose a novel technique for demarcating exogenous $\mathcal{C}(T)$ and endogenous posts $\mathcal{H}(T)$ from a stream of unlabelled messages $\mathcal{U}(T)$ gathered during time $[0, T)$. Additionally, we also set about the task of identifying *extrinsic* users from *organic* users. *Extrinsic* users are the ones who collectively post the majority of exogenous content, while *organic* users mostly discuss the opinions already circulating in the network. Then, based on the categorized posts, we find the optimal parameters $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ by solving a maximum likelihood estimation (MLE) problem. From now onwards, we would write $\mathcal{U}(T)$, $\mathcal{H}(T)$, $\mathcal{C}(T)$ as $\mathcal{U}_T$, $\mathcal{H}_T$ and $\mathcal{C}_T$ to lighten the notations. Furthermore, we denote the organic and extrinsic users as $\mathcal{O}$ and $\mathcal{I}$ respectively, with $\mathcal{I} = \mathcal{V}\backslash\mathcal{O}$.

Now, we sail to design an unsupervised learning algorithm to isolate the endogenous events $\mathcal{H}_T$ and exogenous events $\mathcal{C}_T$ from the stream of unlabeled sentiment messages $\mathcal{U}_T$, which is equivalent to assigning each event $e \in \mathcal{U}_T$ into $\mathcal{H}_T$ or $\mathcal{C}_T$. This is achieved by extracting the set of events that comply with the endogenous dynamics with high confidence that in-turn is indicated by a low variance of estimated parameters. More in detail, given a candidate set of endogenous events $\mathcal{H}_T$ and a candidate set of organic users $\mathcal{O}$, the opinion parameters $\boldsymbol{A}_\mathcal{O} = (\boldsymbol{A}_u)_{u\in\mathcal{O}}, \boldsymbol{\alpha}_\mathcal{O} = (\alpha_u)_{u\in\mathcal{O}}$ can be estimated by maximizing the likelihood $\sum_i \log p(m_{u_i}|x^*_{u_i}(t_i))$, *i.e.*, minimizing the following,

$$\min_{\boldsymbol{A}_\mathcal{O},\boldsymbol{\alpha}_\mathcal{O}} \sum_{\substack{e_i\in\mathcal{H}_T\\ u\in\mathcal{O}}} \sigma^{-2}\Big(m_u(t_i) - \alpha_u - \int_0^{t_i} g(t-s)(\boldsymbol{\zeta}(s)\odot d\boldsymbol{P}(s))^T\boldsymbol{A}_u\Big)^2$$
$$+ c||\boldsymbol{A}_\mathcal{O}||_F^2 + c||\boldsymbol{\alpha}_\mathcal{O}||_2^2. \tag{11}$$

Here, the first term is derived using the Gaussian nature of $p|x_u^*(t)$ and the last two are the regularized terms. The optimal parameters $(\hat{\boldsymbol{A}}_\mathcal{O},\ \hat{\boldsymbol{\alpha}}_\mathcal{O})$ depend on the candidate set of endogenous messages $\mathcal{H}_T$. Moreover, various choices of $\mathcal{O} \subseteq \mathcal{V}$ and $\mathcal{H}_T \subseteq \mathcal{U}_T$ give different $\boldsymbol{A}_\mathcal{O}$ and $\boldsymbol{\alpha}_\mathcal{O}$, with different parameter variance. To this end, we compute the estimation covariance as,

$$\mathbb{\Sigma}(\mathcal{H}_T, \mathcal{O}) := \mathbb{E}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T,\ \boldsymbol{\theta} := \text{vec}([\boldsymbol{A}_\mathcal{O},\ \boldsymbol{\alpha}_\mathcal{O}]). \tag{12}$$

Here the expectation is taken over the noise process induced while getting the message sentiment m_i, from the opinion $x^*_{u_i}(t_i)$ according to distribution $p(m_i|x^*_{u_i}(t_i))$. Prior to going into the selection mechanism of $\mathcal{H}_T$ and $\mathcal{O}$, we first look into the expression of covariance matrix $\mathbb{\Sigma}$ in the Lemma 2. Note that, the inference problem given by Eq. (11) is that of regularized least squares estimation, and so the covariance matrix for the optimal parameters can be derived in a closed form given in the following:

Lemma 2. *For a given endogenous message-set $\mathcal{H}_T$ and organic user-set $\mathcal{O}$,*

$$\mathbb{\Sigma}(\mathcal{H}_T, \mathcal{O}) = \operatorname*{diag}_{u\in\mathcal{O}}\,(c\boldsymbol{I} + \sigma^{-2}\sum_{e_i\in\mathcal{H}_T} \boldsymbol{\phi}_i^u \boldsymbol{\phi}_i^{uT})^{-1} \tag{13}$$

where, $\boldsymbol{\phi}_i^u = \mathbb{1}_{u_i=u}[\int_0^{t_i} g(t-s)\boldsymbol{\zeta}(s)\odot d\boldsymbol{P}(s),\ 1]$. $\mathbb{1}_\mathcal{X}$ is the indicator function with respect to X.

Our objective is to identify $\mathcal{O}$ and $\mathcal{H}_T$, given their sizes $N_\mathcal{O}$ and $N_\mathcal{H}$ respectively, so that $\mathbb{\Sigma}(\mathcal{H}_T, \mathcal{O})$ is small. Such a set of selected users $\mathcal{O}$ and demarcated message-set $\mathcal{H}_T$ would then follow endogenous opinion dynamics more faithfully than their complements $\mathcal{V}\backslash\mathcal{O}$ and $\mathcal{U}_T\backslash\mathcal{H}_T$ respectively. In order to compute the best candidate for $\mathcal{H}_T$ and $\mathcal{O}$, we need to minimize a suitable function $\Omega(\mathcal{H}_T, \mathcal{O})$ which is some measure of $\mathbb{\Sigma}(\mathcal{H}_T, \mathcal{O})$. Now, we define,

$$\Omega(\mathcal{H}_T, \mathcal{O}) := \text{tr}\,[\log\mathbb{\Sigma}(\mathcal{H}_T, \mathcal{O})], \tag{14}$$

where $\log\mathbb{\Sigma}$ is the matrix logarithm of $\mathbb{\Sigma}$. We observe that, $\text{tr}\,[\log\mathbb{\Sigma}(\mathcal{H}_T, \mathcal{O})] = \log\,[\det(\mathbb{\Sigma}(\mathcal{H}_T, \mathcal{O}))]$. Therefore $\Omega(\mathcal{H}_T, \mathcal{O})$ can also be viewed as a complexity measure of $\mathbb{\Sigma}(\mathcal{H}_T, \mathcal{O})$ [47], that makes it a good candidate for minimizing $\mathbb{\Sigma}$. In fact, minimizing $\Omega(\mathcal{H}_T, \mathcal{O})$ is equivalent to minimizing the sum of logarithms of eigenvalues of $\mathbb{\Sigma}$, which would effectively make $\mathbb{\Sigma}(\mathcal{H}_T, \mathcal{O})$ as small as possible. Hence, by defining $f(\mathcal{H}_T, \mathcal{O}) := -\Omega(\mathcal{H}_T, \mathcal{O})$, we pose the following optimization problem to obtain the best cardinality constrained candidate sets $\mathcal{H}_T$, and $\mathcal{O}$:

$$\underset{\mathcal{O}\in\mathcal{V},\ \mathcal{H}_T\in\mathcal{U}_T}{\text{maximize}}\ f(\mathcal{H}_T, \mathcal{O})$$
$$\text{subject to, } |\mathcal{O}| = N_\mathcal{O},\ |\mathcal{H}_T| = N_\mathcal{H} \tag{15}$$

We will rely on a greedy heuristic for maximizing f (Algorithm 1), that, we would show later, gives an $(1-1/e)$ approximation bound. Before going to that, we first specify two properties defined for any multidimensional set function $h(X_1, X_2)$ in general (Definition 3) . We would show that, f specifically enjoys these properties, thereby affording an approximation guarantee from the proposed simple greedy algorithm.

Definition 3. *A multidimensional set function $h(X_1, X_2)$ in two set arguments $X_1 \subseteq U_1$ and $X_2 \subseteq U_2$, with $U_1 \cap U_2 = \emptyset$, is said to be (i)* ***Conditionally submodular (monotone)****, if h is submodular (monotone) in X_1, while keeping X_2 fixed and vice versa; (ii)* ***Jointly***

***submodular**, if for any two sets $X_1 \subseteq \overline{X}_1$ and $X_2 \subseteq \overline{X}_2$, $x \notin \overline{X}_1$ and $y \notin \overline{X}_2$, then $h(X_1 \cup x, X_2 \cup y) - h(X_1, X_2) \geq h(\overline{X}_1 \cup x, \overline{X}_2 \cup y) - h(\overline{X}_1, \overline{X}_2)$, and jointly monotone, if $h(X_1 \cup x, X_2 \cup y) \geq h(X_1, X_2)$.*

Note that, the joint submodularity of $h(X_1, X_2)$ is different from the traditional bisubmodular property which is often encountered in the submodular optimization literature [48–53]. While a bisubmodular function $h(X_1, X_2)$ is defined over two *similar type* of subsets X_1 and X_2 of same universal set U (*i.e.* $X_1, X_2 \subseteq U$), the input sets X_1 and X_2 in Definition 3 are very different subsets selected from two separate universal sets U_1 and U_2. Indeed, in our context too, the set arguments $\mathcal{H}_T$ and O for f of our proposal are drastically different subsets with two unalike universal sets ($\mathcal{U}_T$ and $\mathcal{V}$). Consequently, the existing techniques of bisubmodular optimization cannot be applied to solve (15).

Theorem 4 (**Characterizing** f). *(i) $f(\mathcal{H}_T, O)$ is **conditionally submodular and monotone** in each of $\mathcal{H}_T$ and O. (ii) If $\mathcal{V}(\mathcal{H}_T) \subseteq O$, where $\mathcal{V}(\mathcal{H}_T)$ is the set of users of the message set $\mathcal{H}_T$, then $f(\mathcal{H}_T, O)$ is **jointly submodular and monotone** in both $\mathcal{H}_T$ and O.*

Proof Idea: The key to the proof of (i) relies on mapping the given set-function f to suitably chosen continuous functions $g_1(p)$ and $g_2(p)$, so that, $g_1(1) > g_1(0)$ and $g_2(1) < g_2(0)$ imply the conditional monotonicity and submodularity of f respectively. The rest of the proof of (i) focuses to show that $\frac{d}{dp} g_1(p) > 0$ and $\frac{d}{dp} g_2(p) < 0$ which ensures $g_1(1) > g_1(0)$ and $g_2(0) > g_2(1)$. Such a method is adopted in networked-controllability analysis [54], that is generalized here for more complex networked dynamical systems. The proof of part (ii) follows directly from part (i) of the theorem. A detailed proof is provided in [46].

Note that, Part (ii) of the above theorem operates on an implicit assumption that, $\mathcal{V}(\mathcal{H}_T) \subseteq O$; in words, the users of message set $\mathcal{H}_T$ belong to O. Otherwise, if we suppose $v \in \mathcal{V}(\mathcal{H}_T)$ but $v \notin O$, then the events posted by user v, *i.e.* the vectors $\boldsymbol{\phi}_i^v$ are not contributing to $f(\mathcal{H}_T, O)$. Therefore $f(\mathcal{H}_T \backslash \{e_v\}, O) = f(\mathcal{H}_T, O)$, where e_v is an message posted by user v. So, the assumption specifies a choice for minimal user-set O for $f(\mathcal{H}_T, O)$, and hence is not restrictive.

Maximization of $f(\mathcal{H}_T, O)$**:** Since f is jointly submodular in $\mathcal{H}_T$ and O, its maximization requires further modification of the traditional greedy approach adopted for maximizing submodular function of a single set [55]. The maximization routine is formally shown in Algorithm 1. At each step, it greedily adds event e to $\mathcal{H}_T$ and the user u to O sequentially, by maximizing the marginal gain $f(\mathcal{H}_T \cup \{e\}, O \cup \{u\}) - f(\mathcal{H}_T, O)$ (step 7, Algorithm 1) , until the total number of users reaches N_O (step 5–11). Once $|O|$ hits N_O, it does not add any further user, but keeps choosing events e from $\mathcal{H}_T$, that maximizes $f(\mathcal{H}_T \cup \{e\}, O) - f(\mathcal{H}_T, O)$ until the $\mathcal{H}_T$ reaches $N_{\mathcal{H}}$.
Perhaps surprisingly, the modified greedy algorithm too, achieves a constant $(1 - 1/e)$ fraction of maximum of $f(\mathcal{H}_T, O)$.

Lemma 5 (**Solution-quality**). *Algorithm 1 admits an $(1 - 1/e)$ approximation bound for $f(\mathcal{H}_T, O)$.*

Algorithm 1: Y=CherryPick $(f, N_O, N_{\mathcal{H}}, \mathcal{V}, \mathcal{U}_T)$

1: **Initialization**:
2: $\mathcal{H}_T \leftarrow \emptyset,\ O \leftarrow \emptyset,\ \mathcal{I} \leftarrow \mathcal{V},\ C_T \leftarrow \mathcal{U}_T$
3: **General subroutine**:
4: **while** $|\mathcal{H}_T| < N_{\mathcal{H}}$ **do**
5: **if** $|O| < N_O$ **then**
6: /*Choose e and u in a greedy manner */
7: $(e, u) \leftarrow \arg\max_{e,u} f(\mathcal{H}_T \cup \{e\}, O \cup \{u\}) - f(\mathcal{H}_T, O)$
8: $C_T \leftarrow C_T \backslash \{e\}$, $\mathcal{I} \leftarrow \mathcal{I} \backslash \{u\}$
9: /*Update endogenous message-set and user-set */
10: $\mathcal{H}_T \leftarrow \mathcal{H}_T \cup \{e\},\ O \leftarrow O \cup \{u\}$
11: **end if**
12: /*The user budget is reached. $|O| = N_O$ So, select only messages from now on. */
13: $e \leftarrow \arg\max_{e \in \mathcal{U}_T} f(\mathcal{H}_T \cup \{e\}, O) - f(\mathcal{H}_T, O)$
14: $C_T \leftarrow C_T \backslash \{e\}$
15: /*Update only endogenous message-set */
16: $\mathcal{H}_T \leftarrow \mathcal{H}_T \cup \{e\}$
17: **end while**
18: Y = $(\mathcal{H}_T, O, C_T, \mathcal{I})$
19: **return** Y.

The overview of the proof is similar (yet not identical) to that of ordinary submodular function [56]. The key to the proof is sequentially updating the lower bound of the f obtained in Algorithm 1, using its joint submodularity. Such a lower bound, after a large number of updates, approaches to $(1 - 1/e)$. For the sake of brevity, the proof is omitted here, but is given in [46].
The users other than the selected organic ones $\mathcal{I} = \mathcal{V} \backslash O$ are the extrinsic users who would be actuated for steering the opinion of others during opinion control in Section 5.

Algorithm 2: Parameter Estimation

Input: N_O, $N_{\mathcal{H}}$, $\mathcal{G}$, $\mathcal{U}_T$
Output: $(\boldsymbol{\alpha}^*, \boldsymbol{\mu}^*, \boldsymbol{A}^*, \boldsymbol{B}^*)$
/*First find the endogenous messages */
$(\mathcal{H}_T, O, C_T, \mathcal{I})$=CherryPick $(f, N_O, N_{\mathcal{H}}, \mathcal{V}, \mathcal{U}_T)$
/*Estimate parameters over only $\mathcal{H}_T$ */
$(\boldsymbol{\alpha}^*, \boldsymbol{\mu}^*, \boldsymbol{A}^*, \boldsymbol{B}^*) = \text{argmax}\ \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{A}, \boldsymbol{B} | \mathcal{H}_T, O)$
return $\boldsymbol{\alpha}^*, \boldsymbol{\mu}^*, \boldsymbol{A}^*, \boldsymbol{B}^*$.

The event-set $\mathcal{H}_T$ thus obtained would be used next to estimate all the parameters $\boldsymbol{A}, \boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{B}$ (See Algorithm 2) by maximizing $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{A}, \boldsymbol{B} | \mathcal{H}_T, O)$ which is same as,

$$\sum_{e_i \in \mathcal{H}_T} p(m_{u_i} | x^*_{u_i}(t_i)) + \sum_{e_i \in \mathcal{H}_T} \log(\lambda_{u_i}(t_i)) - \sum_{u \in O} \int_0^T \lambda^*_u(s) ds.$$

Since $\mathcal{L}$ is concave function, one can maximize this efficiently. We adopt the method given by the authors in [6], which can accurately computes the parameters.

5 STEERING OPINION DYNAMICS

In this section, we formally state the online opinion shaping problem, and then we tackle it from the perspective of stochastic control of jump SDEs (Eqs. (16)). First, we modify the endogenous dynamics given by Eq. (10) from control viewpoint.

$$\begin{aligned} d\mathbf{x}^*(t) &= \omega(\boldsymbol{\alpha} - \mathbf{x}^*(t))dt + \mathbf{A}\mathbf{m}(t) \odot d\boldsymbol{N}(t) + \boldsymbol{C} d\boldsymbol{M}^+(t) - \boldsymbol{C} d\boldsymbol{M}^-(t) \\ d\boldsymbol{\lambda}^*(t) &= \nu(\boldsymbol{\mu} - \boldsymbol{\lambda}^*(t))dt + \mathbf{B}\, d\boldsymbol{N}(t) + \mathbf{D} d\boldsymbol{M}(t) \end{aligned} \quad (16)$$

In words, a set of users control the endogenous opinion process $\boldsymbol{x}^*(t)$, by posting positive (+1 opinion) and negative messages (−1 opinion) associated with counting process $\boldsymbol{M}^{+}(t)$ and $\boldsymbol{M}^{-}(t)$. Here $\boldsymbol{C}$ and $\boldsymbol{D}$ are matrices of size $|\mathcal{V}|\times|\mathcal{I}|$. They are submatrices of $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively, induced by the selected control users. That is, $\boldsymbol{C} = \boldsymbol{A}_{\mathcal{V},\mathcal{I}}$ and $\boldsymbol{D} = \boldsymbol{B}_{\mathcal{V},\mathcal{I}}$. Our objective is to find the intensity of $\boldsymbol{\eta}^{\pm}(t)$ of the control counting processes $\boldsymbol{M}^{\pm}(t)$, that optimally steer the opinions of the users in a desired way. Additionally, we assume that $\xi_{\max}(\boldsymbol{B}) << 1$. In reality, we actually found that most datasets satisfy this property, that is the temporal influences take quite small numbers.

5.1 The online opinion shaping problem

Given a directed network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a small set of control users $\mathcal{I}$, we aim to find the optimal control intensity $\boldsymbol{\eta}^{\pm}(t)$ that minimizes the expected value of a particular loss function $\ell(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), \boldsymbol{\eta}^{\pm}(t))$ of the overall endogenous opinions of the network, and the control rates over a time window $(t_0, t_f]$, *i.e.*,

$$\min_{\boldsymbol{\eta}^{\pm}(t)} \mathbb{E}[\phi(\boldsymbol{x}^*(t_f)) + \int_{t_0}^{t_f} \ell(\boldsymbol{x}^*(t), \boldsymbol{\eta}^{\pm}(t))\, dt]. \tag{17}$$

We define $\ell(\boldsymbol{x}^*(t), \boldsymbol{\eta}^{\pm}(t))$ as,

$$\frac{1}{2}(\boldsymbol{x}^*(t) - \boldsymbol{x}^{\text{Track}})^T\, \boldsymbol{Q}\, (\boldsymbol{x}^*(t) - \boldsymbol{x}^{\text{Track}}) + \sum_{s\in\{+1,-1\}} \boldsymbol{\eta}^{sT}(t)\, \boldsymbol{S}\, \boldsymbol{\eta}^{s}(t)$$

$$\& \;\; \phi(\boldsymbol{x}^*(t_f)) = \frac{1}{2}(\boldsymbol{x}^*(t_f) - \boldsymbol{x}^{\text{Track}})^T\, \boldsymbol{F}\, (\boldsymbol{x}^*(t_f) - \boldsymbol{x}^{\text{Track}}). \tag{18}$$

Here $\boldsymbol{x}^{\text{Track}}$ is the desired opinion-vector to which the controller aims to steer. Furthermore, using penalty term for $\boldsymbol{\eta}^{\pm}(t)$ in the expression of $\ell(\boldsymbol{x}^*(t), \boldsymbol{\eta}^{\pm}(t))$, we will limit the number of posts we steer. Here $\boldsymbol{Q}$, $\boldsymbol{F}$ and $\boldsymbol{S}$ are p.s.d matrices with $\boldsymbol{Q}_{ij} \geq 0$, $\boldsymbol{F}_{ij} \geq 0$ and $\boldsymbol{S}_{ij} \geq 0$ for all $i, j \in [n]$.

5.2 Stochastic optimal control (SOC) algorithm

At the very outset, our aim is to compute $\boldsymbol{\eta}^{\pm}(t)$, by minimizing the loss proposed above (Eq. (18)). To this aim, we first define an optimal cost-to-go function J and then, using the Bellman's principle of optimality [57], we derive and finally solve the Hamilton-Jacobi-Bellman (HJB) equation to find the optimal control intensity.

While, solving such an SOC often follows the standard roadmap [58] adopted here, the challenges across different milestones are quite application specific, and therefore lacks a unified solution proposal. Hence, in the context of our problem, we follow a similar direction, tackle the difficulties at different steps, and finally provides a novel closed form expression of the cost to go J that in turn is used to compute $\boldsymbol{\eta}^{\pm}(t)$.

The optimal cost-to-go $J(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), t)$ is defined as,

$$J(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), t) = \min_{\boldsymbol{\eta}^{\pm}(t, t_f]} \mathbb{E}[\phi(\boldsymbol{x}^*(t_f)) + \int_{t}^{t_f} \ell(\boldsymbol{x}^*(t), \boldsymbol{\eta}^{\pm}(t))\, dt] \tag{19}$$

which is the minimum of the expected cost value of going from the state $(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t))$ at time t to the final state at time t_f,

The value of this cost-to-go function $J(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), t)$ would be used to find the optimal control $\boldsymbol{\eta}^{\pm}(t)$. Therefore we set about for accurate estimation of J. To this aim, we first find the differential expression dJ using Bellman's principle of optimality [57, 58].

Theorem 6 (Differential HJB equation). *[58] The optimal cost-to-go function defined by Eq. 19 satisfies the following differential equation:*

$$\min_{\boldsymbol{\eta}^{\pm}(t, t+dt]} \{\mathbb{E}\left[dJ(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), t)\right] + \ell(\boldsymbol{x}^*(t), \boldsymbol{\eta}^{\pm}(t))dt\} = 0. \tag{20}$$

In order to solve Eq. (20), we expand the differential dJ by using chain rule, *i.e.* differentiating w.r.t all the variables $\boldsymbol{x}^*(t)$, $\boldsymbol{\lambda}^*(t)$ and t, and take into account of the jump process $d\boldsymbol{N}(t)$ and $d\boldsymbol{M}(t)$. Finally we have:

Lemma 7. *The HJB equation given by Eq. 20 on expanding satisfies the following:*

$$\begin{aligned}\min_{\boldsymbol{\eta}^{\pm}(t,t+dt]} \Big\{ & \frac{\partial J}{\partial t} - \omega(\boldsymbol{x}^*(t) - \boldsymbol{\alpha})^T \frac{\partial J}{\partial \boldsymbol{x}^*} - \nu(\boldsymbol{\lambda}^*(t) - \boldsymbol{\beta})^T \frac{\partial J}{\partial \boldsymbol{\lambda}^*} \\ & + \sum_{i\in\mathcal{V}} \lambda_i^*(t)\mathbb{E}\Delta_{\boldsymbol{A}_i m_i, \boldsymbol{B}_i} J + \sum_{i\in\mathcal{I}} \left[\boldsymbol{\eta}_i^{+}(t)\Delta_{\boldsymbol{A}_i,\boldsymbol{B}_i} J + \boldsymbol{\eta}_i^{-}(t)\Delta_{-\boldsymbol{A}_i,\boldsymbol{B}_i} J\right] \\ & + \ell(\boldsymbol{x}^*(t), \boldsymbol{\eta}^{\pm}(t))\Big\} = 0 \end{aligned} \tag{21}$$

where $\Delta_{\boldsymbol{a},\boldsymbol{b}} J$ is given by $\Delta_{\boldsymbol{a},\boldsymbol{b}} J = J(\boldsymbol{x}^*(t)+\boldsymbol{a}, \boldsymbol{\lambda}^*(t)+\boldsymbol{b}, t) - J(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), t)$ and the expectation is taken over the noise induced due to sampling of m_i from $x_i^*(t)$(Proved in [46]).

If we minimize over $\boldsymbol{\eta}^{\pm}$ on the LHS of the HJB equation in Eq. 21, we have $\boldsymbol{\eta}^{\pm} = -\boldsymbol{S}^{-1}\Delta_{\pm\boldsymbol{A},\boldsymbol{B}} J$. Here $\Delta_{\pm\boldsymbol{A},\boldsymbol{B}}$ is a vector whose i^{th} element is $\Delta_{\pm\boldsymbol{A}_i,\boldsymbol{B}_i}$ which is defined in the context of Lemma 7. Substituting these optimal values of $\boldsymbol{\eta}^{\pm}$ in the Eq. 21, we have:

$$\begin{aligned} 0 = & \frac{\partial J}{\partial t} - \omega(\boldsymbol{x}^*(t) - \boldsymbol{\alpha})^T \frac{\partial J}{\partial \boldsymbol{x}^*} - \nu(\boldsymbol{\lambda}^*(t) - \boldsymbol{\beta})^T \frac{\partial J}{\partial \boldsymbol{\lambda}^*} \\ & + \sum_{i\in\mathcal{V}} \lambda_i^*(t)\mathbb{E}\Delta_{\boldsymbol{A}_i m_i, \boldsymbol{B}_i} J - \frac{1}{2}\sum_{s\in\pm} \Delta_{s\boldsymbol{A},\boldsymbol{B}} J^T \boldsymbol{S}^{-1} \Delta_{s\boldsymbol{A},\boldsymbol{B}} J \\ & + \frac{1}{2}(\boldsymbol{x}^*(t) - \boldsymbol{x}^{\text{Track}})^T\, \boldsymbol{Q}\, (\boldsymbol{x}^*(t) - \boldsymbol{x}^{\text{Track}}) \end{aligned} \tag{22}$$

with $J(\boldsymbol{x}^*(t_f), \boldsymbol{\lambda}^*(t_f), t_f) = \phi(\boldsymbol{x}^*(t_f))$ as the terminal condition. Finally we reach the optimal solution of $J(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), t)$ which is given by the following:

Lemma 8. *The solution of the nonlinear differential equation given by Eq. 22 in the space of all polynomials, is the following quadratic form (Proved in [46]):*

$$\begin{aligned} J(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), t) = & \; h(t) + \boldsymbol{g}(t)^T \boldsymbol{\lambda}^*(t) + \boldsymbol{f}^T(t)\boldsymbol{x}^*(t) \\ & + \boldsymbol{\lambda}^{*T}(t)\boldsymbol{V}(t)\boldsymbol{x}^*(t) + \frac{1}{2}\boldsymbol{\lambda}^{*T}(t)\boldsymbol{K}(t)\boldsymbol{\lambda}^*(t) + \frac{1}{2}\boldsymbol{x}^{*T}(t)\boldsymbol{H}(t)\boldsymbol{x}^*(t) \end{aligned} \tag{23}$$

where the coefficient-tuple $\Pi(t) = (\boldsymbol{h}, \boldsymbol{g}, \boldsymbol{f}, \boldsymbol{V}, \boldsymbol{K}, \boldsymbol{H})(t)$ can be found by solving a set of six differential equations. For brevity the exact form of differential equations is given in [46].

$$\dot{\boldsymbol{\Pi}}(t) = \textbf{Riccati}(\boldsymbol{\Pi}(t)) \tag{24}$$

This matrix Riccati differential equation, can be solved using many well-known efficient numerical solvers [59]. Finally, given the above J, we compute the optimal intensity as,

$$\boldsymbol{\eta}^{\pm} = -\boldsymbol{S}^{-1}\Delta_{\pm\boldsymbol{A},\boldsymbol{B}} J \tag{25}$$

which was derived using the differential HJB equation.

Dataset	Mean squared error						Failure rate					
	CherryPick	SLANT	AsLM	DeGroot	Voter	B-Voter	CherryPick	SLANT	AsLM	DeGroot	Voter	B-Voter
Bollywood	0.104 (33.6)	0.157	0.667	0.752	0.889	0.693	0.043 (41.0)	0.072	0.517	0.584	0.550	0.584
Series	0.110 (48.2)	0.213	0.611	0.634	0.836	0.642	0.097 (23.2)	0.125	0.481	0.509	0.527	0.513
Soccer	0.090 (31.9)	0.132	0.543	0.588	1.201	0.702	0.028 (53.4)	0.061	0.427	0.449	0.452	0.452
Verdict	0.060 (33.3)	0.090	0.598	0.685	1.000	1.081	0.057 (22.3)	0.073	0.452	0.477	0.465	0.475
Elections	0.146 (24.1)	0.193	0.510	0.616	1.260	0.701	0.073 (26.1)	0.098	0.348	0.404	0.349	0.366

Table 1: Forecasting performance across all the models using five datasets and 2 error metrics for $|\mathcal{H}_T| = 0.8|\mathcal{U}_T|$ and $T_f = 4$ hrs. The yellow (cyan) cells reflect the best (second best) predictor. Numbers in the brackets denote percentage improvement over the nearest baseline.

Dataset	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{U}(T)\rvert$	$\mathbb{E}[\mathbf{m}]$	$\mathbf{std}[m]$
Bollywood	1031	34952	46845	0.5101	0.2310
Series	947	10253	13203	-0.0216	0.3177
Soccer	703	4154	8319	0.1779	0.1521
Verdict	1059	17452	9950	0.5170	0.1870
Elections	533	20067	18704	-0.0186	0.7135

Table 2: Real datasets statistics.

6 EXPERIMENTS

We provide a comprehensive evaluation of CherryPick from both modeling and shaping viewpoints, across the following real datasets (also summarized in Table 2) corresponding to various real-world events, collected from Twitter.

1. **Bollywood:** Verdict that declared guilty to Salman Khan (a popular Bollywood movie star) for causing death of a person by rash and negligible driving, from May 5 to May 16, 2015.
2. **Series:** The promotion on the TV show "Games of Thrones", from May 4 to May 12, 2015.
3. **Soccer:** Champions League final in 2015, between Juventus and Real Madrid, from May 8 to May 16, 2015.
4. **Verdict:** Verdict for the corruption-case against Jayalalitha, an Indian politician, from May 6 to May 17, 2015..
5. **Elections:** Presidential election in the United-States, from April 7 to 13, 2016.

For all datasets, we follow a very standard setup for both network construction and message sentiment computation [5, 6, 21]. We built the follower-followee network for the users that posted related tweets using the Twitter rest API[1]. Then, we filtered out users that posted less than 200 tweets during the account lifetime, follow less than 100 users, or have less than 50 followers. For each dataset, we compute the sentiment values of the messages using a popular sentiment analysis toolbox [35]. Here, the sentiment takes values $m \in [-1, 1]$ and we consider the sentiment polarity to be simply $\text{sign}(m)$. Note that, while other sentiment analysis tools [34] can be used to extract sentiments from tweets, we appeal to [35] due to two major reasons– its ability of accurately extracting sentiments from short informal texts like tweets, and its wide usage in validating data-driven opinion models [5, 6].

6.1 Effect of CherryPick on opinion modeling

We evaluate the efficacy of demarcation of endogenous and exogenous messages, by measuring the predictive prowess of the associated opinion model given by Eq. (4), in comparison with five state-of-the-art opinion models, e.g. SLANT [6], the asynchronous linear model (AsLM) [5], DeGroot's model [2], the Voter model [3] and the Biased Voter model [4].

Evaluation protocol and metrics: Given a temporal stream of sentiment messages $\mathcal{U}$, we first split it into training and test set, where training set consists of first 90% of the total number of messages. We first demarcate these messages $\mathcal{U}_T$, say collected until time T, into endogenous $\mathcal{H}_T$ and exogenous messages C_T, and then estimate the parameters over the classified $\mathcal{H}_T$. During categorization, we took a range of values of pre-specified value of $|\mathcal{H}_T|$ ($N_{\mathcal{H}}$), the pre-specified number of organic messages. However, we assumed $O = \mathcal{V}$ to extract the endogenous dynamics from all users (See Eq. (15)). Note here we only assess the benefits of message classification proposal; the efficacy of user classification is discussed in the following subsection (Sec. 6.2). Finally, using this estimated model, we forecast the sentiment value m for each message in the test set given the history up to T_f hours before the time of the message as $\hat{m} = E_{\mathcal{H}_t \setminus \mathcal{H}_{t-T_f}}[x^*_u(t)|\mathcal{H}_{t-T_f}]$ that we compute using an efficient simulation method given by [6, 60]. For predicting opinions using discrete models e.g. AsLM, DeGroot, Voter, and Biased Voter, which operate in discrete time, we run N_{T_f} rounds of simulation in $(t - T_f, t)$, where N_{T_f} is the number of posts during this interval. We measure the performance of our model along with the baselines, in terms of: (i) the *Mean squared error (MSE)* between the actual and the estimated sentiment value, *i.e.*, $\mathbb{E}[(m - \hat{m})^2]$, and (ii) the *Failure rate (FR)* which is the polarity prediction error, *i.e.*, $\mathbb{E}[\mathbb{1}_{\text{sign}(m)\neq\text{sign}(\hat{m})}]$.

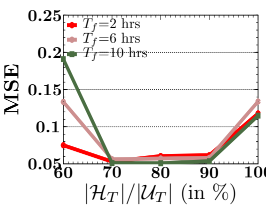
Comparison with baselines: Table 1 dissects a comparative sketch of the prediction error of five state-of-the-art methods and our proposal for $|\mathcal{H}_T| = 0.8|\mathcal{U}_T|$, and $T_f = 4$ hours. The left half of the table reports Mean square error, while the rest reports Failure rate. We observe that, CherryPick offers a significant performance boost in comparison to all its competitors, including SLANT, an immediate counterpart of our proposal, which however, does not model exogenous signals.

Voter model and its variants [3, 4]: The performance of Voter model and Biased Voter model are quite poor. Voter model allows a user to form her opinion, by randomly selecting opinion from one of her neighbors. Such an update strategy keeps the set of opinions invariant throughout the diffusion process. Hence, it operates in closed social network, as opposed to the reality where social networks are open system allowing signals to flow in and out. Biased Voter model aims to overcome some limitations of Voter model by introducing node weights. However it also ignores the effect of externalities, and it fares poorly than most other baselines.

[1] https://dev.twitter.com/rest/public

Class	Example tweets
Exogenous (18:08:51)	#BREAKING: Donald Trump campaign manager will not be prosecuted. A total joke. : M_1
Endogenous (18:18:17)	Sad! Donald Trump, quit your whining.: M_2
Endogenous (18:32:09)	Trump is snake oil businessman.: M_3
Endogenous (18:44:20)	Yes, Confirmed! Trump campaign manager won't be Prosecuted. Such a huge disaster. : M_4
Exogenous (18:59:36)	Trump shows farsightedness by shaking hands with Russia. Liberals will never understand.: M_5

Table 3: Anecdotal examples for accurate message demarcation using CherryPick on Elections dataset. Time in bracket indicates the posting time.



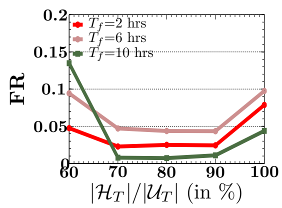


Figure 1: Performance change with $|\mathcal{H}_T|$ for Verdict dataset.

Linear models [2, 5]: The performance of linear models (AsLM and DeGroot) are better than the Voter models. These models attempt to capture the endogenous opinion dynamics via the edge-weights. However, they ignore the effect of posting time of messages on opinion diffusion, which constrains their predictive power.

SLANT [6]: Among all the baselines, we observe that SLANT is the best predictor. As opposed to the other models which do not consider the role of temporal dynamics, SLANT incorporates the influence of past messages on the opinion diffusion process. Furthermore, using the constant term $\boldsymbol{\alpha}$, it often aims to capture the overall effect of external signals, but succeeds only partially. Since it assumes all the collected messages are endogenous, it fails to probe the exogenous signals at individual user and message level, leaving a substantial space for improvement.

CherryPick: CherryPick accurately captures the effect of temporal dynamics of historical data, as well as incorporates the effect of exogenous signals at the message and user level. Its principled demarcation paradigm aids to identify the endogenous messages. In contrast to SLANT and other competitors, the exogenous messages are no more modeled endogenously in this case. Using such a refined set of training data, the parameters are inferred more accurately, and as a result the estimated model precisely brings out the complex opinion dynamics, thereby showing a substantial performance boost in comparison to its competitors. In fact, due to the dropped-out exogenous messages, CherryPick effectively uses a smaller amount of training data than its competitors, and yet outperforms them in terms of forecasting ability. This is because, the final training set used in parameter estimation is less noisy, and contains only the endogenous signals, thereby aiding in a better inference.

Performance variation with $|\mathcal{H}_T|$: Figure 1 describes the variation of forecasting performance for different values of $|\mathcal{H}_T|$, across two representative datasets. We observe that, upon increasing $|\mathcal{H}_T|$, the prediction error first decreases and then increases, strongly indicating an optimum number of endogenous messages in the temporal data. A small value of $|\mathcal{H}_T|$, overestimates the effect of exogenous signal, while a large value of $|\mathcal{H}_T|$ ignores its effect. A well-calibrated value for $|\mathcal{H}_T|$ optimally selects an appropriate set of messages, which helps in accurate opinion forecasting. Furthermore, we observe this optimal value of $|\mathcal{H}_T|$ remains almost consistent throughout a moderate variation of T_f, which reflects a substantial robustness of our proposal.

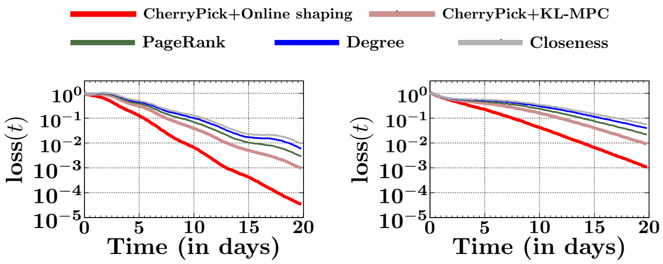
Illustration with examples. Table 3 shows a few example messages from a conversation around US election (Elections dataset), that CherryPick has successfully categorized. The conversation largely reflects negative or anti-trump sentiments (M_1 to M_4). We observe that, M_1 which is triggered by externalities, subsequently influences the generation of endogenous messages M_2 to M_4. Note that, despite having a strong content similarity with M_1, M_4 is not exogenous, since it is generated from the process of opinion formation internal to the system. Finally, we notice that M_5 contains a positive opinion, and the node positing it seems not to be influenced by posts around its neighborhood rather has imported a new tweet (opinion) from 'outside'. CherryPick correctly spots it and labels as exogenous.

6.2 Effect of CherryPick on opinion shaping

The performance of our proposed opinion control policy depends on two factors: (i) the prudence shown by CherryPick in selecting control users $\mathcal{I}$, and (ii) the efficiency of the online opinion shaping framework (Sec 5) – both of them are evaluated in this section. Since, we observe that most (> 90%) of the users in all the datasets have positive initial opinions ($\alpha_u > 0$), choosing a loss function to steer opinion of each user to an extreme negative opinion, should appropriately test the shaping performance, and therefore, we set, $\boldsymbol{x}^{\text{Track}} = -\mathbf{1}$ (Eq. (18)). To minimize the corresponding loss function, at each timestamp, the shaping algorithm supplies suitable pairs of the exogenous intensities $\boldsymbol{\eta}^{\pm}(t)$ of the control users, which in turn regulate the number of control messages ($\boldsymbol{M}^{\pm}(t)$) , for guiding the opinion dynamics to reach the desired state. In this context, we define $\overline{\boldsymbol{M}}^{\pm}(t_f) = \sum_{u\in\mathcal{V}} M_u^{\pm}(t_f)$ and $\overline{\boldsymbol{M}}(t_f) = \overline{\boldsymbol{M}}^{+}(t_f) + \overline{\boldsymbol{M}}^{-}(t_f)$.

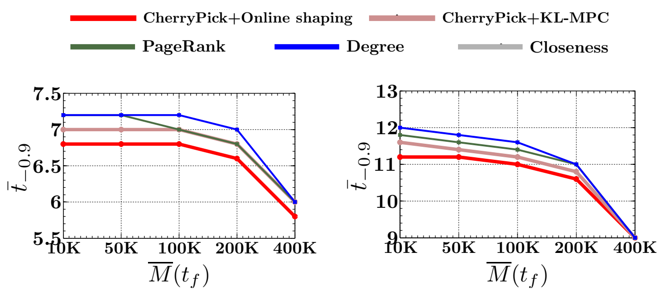
Baseline setup: We compare our proposal with KL-MPC proposed by Wang *et al.* [21], and three centrality based measures - PageRank, Degree and Closeness. The operational principle of the proposed strategy relies on accurate supervision of the number of exogenous posts for both positive and negative opinions. A fair competition between the shaping proposals would, therefore, require a uniformity in the total number of control messages $\overline{\boldsymbol{M}}(t_f)$ across all the baselines. To this aim, we tune the parameters of KL-MPC, so that the corresponding number of control messages approximately matches with the value of $\overline{\boldsymbol{M}}(t_f)$ obtained using our proposal. Furthermore, to understand the efficacy of the online algorithm, we used the same set of extrinsic users $\mathcal{I}$ as control users in KL-MPC. The centrality based measures, on the other hand, distributes the intensities $\boldsymbol{\eta}^{\pm}(t)$ proportionally with the users' centrality scores in the network, thereby triggering same number of control messages across time.

Metrics: We compare the performance using two measures: (i) $\mathbf{loss}(t) := ||\boldsymbol{x}^*(t) - \boldsymbol{x}^{\text{Track}}||^2$, *i.e.* the tracking error indicating how far is the current opinion from the target vector, and (ii) *latency*,



(a) Bollywood (b) Series

Figure 2: Temporal variation of opinion shaping performance for all the baselines, in terms of loss(t), with $|\mathcal{O}|= 0.8|\mathcal{V}|$ (*i.e.* $|\mathcal{I}|= 0.2|\mathcal{V}|$) and $\overline{M}(t_f) \approx 200K$, across two representative datasets.



(a) Soccer (b) Elections

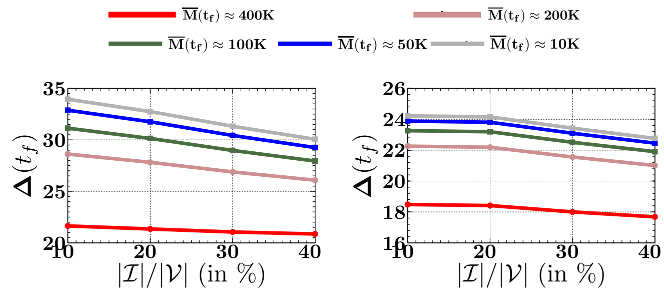
Figure 3: Variation of latency i.e. the average time $\bar{t}_{-0.9}$ required to reach a milestone opinion -0.9, against the number of control actions, for Soccer and Elections datasets, given $|\mathcal{O}|= 0.8|\mathcal{V}|$ (*i.e.* $|\mathcal{I}|= 0.2|\mathcal{V}|$).

denoted as $t_{-0.9}$, *i.e.* the time at which the mean opinion reaches -0.9 (*i.e.* 90% of target $\mathbf{x}^{\text{Track}}$). and (iii) $\Delta(t_f) := \int_0^{t_f} \mathbf{loss}(s)ds$, area under the error curve.

Comparison with baselines.

Loss(t): Figure 2 gives a comparative analysis of our proposal with the other baselines in terms of **loss**(t), across two representative datasets. We observe that our method consistently outperforms the baselines. We also observe that the centrality based measures fare quite poorly. They assign the control power to the users heuristically and linearly, which renders them ineffective in opinion shaping. To some extent, the poor performance of these measures reveals that the structural properties alone are not very effective measures of influence in the context of opinion dynamics. The performance of KL-MPC is better than the centrality based measures. It is more principled, but gives an open loop and approximate solution. Hence, even after supplying it with high-quality control users (through CherryPick, which the original proposal does not), it performs poorly than the closed-loop online control. Our proposal unifies the user classification, and an optimally designed, closed loop online shaping algorithm in a principled way. CherryPick not only helps it to bring out the extrinsic users, but also offers a soft measure of influence in the context of opinion dynamics - thus emphatically establishing the superiority of our algorithm.

Latency: We next evaluate the performance our algorithm against the baselines, with respect to the total number of control messages. To do that, we compare latency, which is the time $t_{-0.9}$ taken to reach a milestone of opinion -0.9, against $\overline{M}(t_f)$, the number of control actions. Figure 3 describes the results, which shows that the proposed shaping method consistently reaches the milestone opinion faster than its competitors. Furthermore, our proposal shows a greater benefit at low budgets, i.e. it can efficiently steer the opinion dynamics even with a small number of messages.



(a) Verdict (b) Soccer

Figure 4: Variation of $\Delta(t_f)$ with $|\mathcal{I}|$, for our online shaping method, across Verdict and Soccer datasets.

Performance variation with $|\mathcal{I}|$. Figure 4 summarizes the results for variation of cumulative loss $\Delta(t_f)$, with the pre-selected user size $|\mathcal{I}|$. We observe that, as the budget increases, the control performance does not vary much with the number of users. From a practical viewpoint, often difficult challenges are associated with high budgets and high latency. In such a shaping problem, a small selected (through CherryPick) number of control users can steer opinions of the users as effectively as a larger number of not-so effective control users, thus highlighting the power of CherryPick.

7 CONCLUSION

The principal contribution of this paper lies in emphatically establishing the dual nature of message flow over online social network: injection of exogenous opinions and influence-based dynamics, internal to the network. The realization helps us to propose CherryPick, a novel learning methodology to demarcate endogenous and exogenous opinions, identify organic and extrinsic users, and finally illustrate its utility from both opinion modeling and shaping perspective. To this aim, we formulated the user and message classification problem as a joint submodular optimization task in the set of users and messages, which we solved using an efficient greedy algorithm. Furthermore, in order to demonstrate the efficacy of user selection task, we developed the opinion shaping problem in a novel framework of stochastic optimal control, that outputs the intensities with which the selected users should post bipolar opinions, to steer the opinion dynamics in a favorable manner. Finally, on five datasets crawled from Twitter, we showed that our proposal consistently outperforms the existing algorithms in terms of both predictive and shaping prowess. The superior performance is even more remarkable considering the fact that we train our system on smaller (but relevant) amount of data than all competing models. We believe the framework developed here can be effectively used to understand several related traits observed on OSN like 'manufactured' trending topics, which would be our one of the immediate future endeveours.

Acknowledgement: This work was partially supported by Google India PhD Fellowship for Social Computing, DST project SB/S3/EECE /0249/2014.

REFERENCES

[1] P. Clifford and A. Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, 1973.

[2] M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[3] M. E. Yildiz, R. Pagliari, A. Ozdaglar, and A. Scaglione. Voting models in random networks. In *Information Theory and Applications Workshop*, pages 1–7, 2010.

[4] A. Das, S. Gollapudi, and K. Munagala. Modeling opinion dynamics in social networks. In *WSDM*, 2014.

[5] A. De, S. Bhattacharya, P. Bhattacharya, N. Ganguly, and S. Chakrabarti. Learning a linear influence model from transient opinion dynamics. In *CIKM*, 2014.

[6] Abir De, Isabel Valera, Niloy Ganguly, Sourangshu Bhattacharya, and Manuel Gomez Rodriguez. Learning and forecasting opinion dynamics in social networks. In *NIPS*. 2016.

[7] Jan Christian Dittmer. Consensus formation under bounded confidence. *Nonlinear Analysis: Theory, Methods & Applications*, 47(7):4615–4621, 2001.

[8] Igor Douven and Alexander Riegler. Extending the hegselmann–krause model i. *Logic Journal of IGPL*, 18(2):323–335, 2009.

[9] Claudio Altafini and Gabriele Lini. Predictable dynamics of opinion forming for networks with antagonistic interactions. *IEEE Transactions on Automatic Control*, 60(2):342–357, 2015.

[10] Irinel-Constantin Mor, Antoine Girard, et al. Opinion dynamics with decaying confidence: Application to community detection in graphs. *IEEE Transactions on Automatic Control*, 56(8):1862–1873, 2011.

[11] Seyed Rasoul Etesami and Tamer Başar. Game-theoretic analysis of the hegselmann-krause model for opinion dynamics in finite dimensions. *IEEE Transactions on Automatic Control*, 60(7):1886–1897, 2015.

[12] Vincent D Blondel, Julien M Hendrickx, and John N Tsitsiklis. On krause's multi-agent consensus model with state-dependent connectivity. *IEEE transactions on Automatic Control*, 54(11):2586–2597, 2009.

[13] Noah E Friedkin. The problem of social control and coordination of complex systems in sociology: A look at the community cleavage problem. *IEEE Control Systems*, 35(3):40–51, 2015.

[14] Irinel-Constantin Morărescu, Samuel Martin, Antoine Girard, and Aurélie Muller-Gueudin. Coordination in networks of linear impulsive agents. *IEEE Transactions on Automatic Control*, 61(9):2402–2415, 2016.

[15] P. Holme and M. E. Newman. Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review E*, 74(5):056108, 2006.

[16] Przemyslaw A Grabowicz, Niloy Ganguly, and Krishna P Gummadi. Distinguishing between topical and non-topical information diffusion mechanisms in social media. In *ICWSM*, pages 151–160, 2016.

[17] Seth A Myers, Chenguang Zhu, and Jure Leskovec. Information diffusion and external influence in networks. In *KDD*, 2012.

[18] Aris Anagnostopoulos, Ravi Kumar, and Mohammad Mahdian. Influence and correlation in social networks. In *KDD*, 2008.

[19] Trisha Greenhalgh, Glenn Robert, Fraser Macfarlane, Paul Bate, and Olivia Kyriakidou. Diffusion of innovations in service organizations: systematic review and recommendations. *The Milbank Quarterly*, 82(4):581–629, 2004.

[20] Thomas W Valente. Social network thresholds in the diffusion of innovations. *Social networks*, 18(1):69–89, 1996.

[21] Yichen Wang, Grady Williams, Evangelos Theodorou, and Le Song. Variational policy for guiding point processes. *arXiv preprint arXiv:1701.08585*, 2017.

[22] A. Zarezade, U. Upadhyay, H. Rabiee, and M. Gomez-Rodriguez. Redqueen: An online algorithm for smart broadcasting in social networks. In *WSDM '17*, 2017.

[23] Ali Zarezade, Abir De, Hamid Rabiee, and Manuel Gomez-Rodriguez. Cheshire: An online algorithm for activity maximization in social networks. In *arXiv preprint arXiv:1703.02059*, 2017.

[24] E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione. Binary opinion dynamics with stubborn agents. *ACM Transactions on Economics and Computation*, 1(4):19, 2013.

[25] Frank Schweitzer and Laxmidhar Behera. Nonlinear voter models: the transition from invasion to coexistence. *The European Physical Journal B-Condensed Matter and Complex Systems*, 67(3):301–318, 2009.

[26] Federico Vazquez, Paul L Krapivsky, and Sidney Redner. Constrained opinion dynamics: Freezing and slow evolution. *Journal of Physics A: Mathematical and General*, 36(3):L61, 2003.

[27] Xavier Castelló, Víctor M Eguíluz, and Maxi San Miguel. Ordering dynamics with two non-excluding options: bilingualism in language competition. *New Journal of Physics*, 8(12):308, 2006.

[28] Nazareno GF Medeiros, Ana TC Silva, and FG Brady Moreira. Domain motion in the voter model with noise. *Physical Review E*, 73(4):046120, 2006.

[29] Rick Durrett and Simon Levin. Spatial models for species-area curves. *Journal of Theoretical Biology*, 179(2):119–127, 1996.

[30] R. Hegselmann and U. Krause. Opinion dynamics & bounded confidence models, analysis, & simulation. *Journal of Artificial Societies & Social Simulation, '02*, 5(3).

[31] Ulrich Krause. A discrete nonlinear and non-autonomous model of consensus formation. *Communications in difference equations*, pages 227–236, 2000.

[32] Bing Liu. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167, 2012.

[33] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

[34] James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001, 2001.

[35] Aniko Hannak, Eric Anderson, Lisa Feldman Barrett, Sune Lehmann, Alan Mislove, and Mirek Riedewald. Tweetin'in the rain: Exploring societal-scale effects of weather on mood. In *ICWSM*, 2012.

[36] Huayi Li, Zhiyuan Chen, Arjun Mukherjee, Bing Liu, and Jidong Shao. Analyzing and detecting opinion spam on a large-scale dataset via temporal and spatial patterns. In *ICWSM*, pages 634–637, 2015.

[37] Santosh KC and Arjun Mukherjee. On the temporal dynamics of opinion spamming: Case studies on yelp. In *WWW*, pages 369–379, 2016.

[38] Huayi Li, Geli Fei, Shuai Wang, Bing Liu, Weixiang Shao, Arjun Mukherjee, and Jidong Shao. Bimodal distribution and co-bursting in review spam detection. In *WWW*, 2017.

[39] Bidisha Samanta, Abir De, Abhijnan Chakraborty, and Niloy Ganguly. Lmpp: a large margin point process combining reinforcement and competition for modeling hashtag popularity. In *IJCAI*, pages 2679–2685, 2017.

[40] Bhushan Kulkarni, Sumit Agarwal, Abir De, Sourangshu Bhattacharya, and Niloy Ganguly. SLANT+: A nonlinear model for opinion dynamics in social networks. In *ICDM*, 2017.

[41] Bidisha Samanta, Abir De, and Niloy Ganguly. Strm: A sister tweet reinforcement process for modeling hashtag popularity. In *INFOCOM 2017*, pages 1–9. IEEE, 2017.

[42] Krunal Parmar, Samuel Bushi, Sourangshu Bhattacharya, and Surender Kumar. Forecasting ad-impressions on online retail websites using non-homogeneous hawkes processes. In *CIKM*, 2017.

[43] Abir De, Sourangshu Bhattacharya, Sourav Sarkar, Niloy Ganguly, and Soumen Chakrabarti. Discriminative link prediction using local, community, and global signals. *IEEE Transactions on Knowledge and Data Engineering*, 28(8):2057–2070, 2016.

[44] Abir De, Niloy Ganguly, and Soumen Chakrabarti. Discriminative link prediction using local links, node features and community structure. In *ICDM*, 2013.

[45] Abir De, Maunendra Sankar Desarkar, Niloy Ganguly, and Pabitra Mitra. Local learning of item dissimilarity using content and link structure. In *ACM RecSys*, 2012.

[46] Supplementary material. http://www.cnergres.iitkgp.ac.in/supp.pdf.

[47] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[48] Ajit Singh, Andrew Guillory, and Jeff Bilmes. On bisubmodular maximization. In *AISTATS*, 2012.

[49] Anna Huber and Vladimir Kolmogorov. Towards minimizing k-submodular functions. *ISCO*, 7422:451–462, 2012.

[50] Kazutoshi Ando, Satoru Fujishige, and Takeshi Naitoh. A characterization of bisubmodular functions. *Discrete Mathematics*, 148(1-3):299–303, 1996.

[51] Anna Huber, Andrei Krokhin, and Robert Powell. Skew bisubmodularity and valued csps. *SIAM Journal on Computing*, 43(3):1064–1084, 2014.

[52] Satoru Fujishige, Shin-ichi Tanigawa, and Yuichi Yoshida. Generalized skew bisubmodularity: A characterization and a min–max theorem. *Discrete Optimization*, 12:1–9, 2014.

[53] Justin Ward and Stanislav Živný. Maximizing k-submodular functions and beyond. *ACM Transactions on Algorithms*, 12(4):47, 2016.

[54] Tyler H Summers, Fabrizio L Cortesi, and John Lygeros. On submodularity and controllability in complex dynamical networks. *IEEE Transactions on Control of Network Systems*, 3(1):91–101, 2016.

[55] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions–i. *Mathematical Programming*, 14(1):265–294, 1978.

[56] Andreas Krause and Daniel Golovin. Submodular function maximization., 2014.

[57] D. P. Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena Scientific Belmont, MA, 1995.

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

[59] C. K. Garrett. *Numerical integration of matrix Riccati differential equations with solution singularities.* PhD thesis, The University of Texas at Arlington, May 2013.

[60] M. Farajtabar, Y. Wang, M. Gomez-Rodriguez, S. Li, H. Zha, and L. Song. Coevolve: A joint point process model for information diffusion and network co-evolution. In *NIPS*, 2015.

Supplementary material

A Proof-sketch of Proposition 1:

We rewrite Eq. (6) and Eq. (7):

$$\boldsymbol{x}^*(t) = \boldsymbol{\alpha} + \int_0^t g(t-s)\boldsymbol{A}[\boldsymbol{m}(s) \odot d\boldsymbol{N}(s) + \boldsymbol{w}(s) \odot d\boldsymbol{M}(s)] \tag{26}$$

$$\boldsymbol{\lambda}^*(t) = \boldsymbol{\mu} + \int_0^t \boldsymbol{B}\kappa(t-s)[d\boldsymbol{N}(s) + d\boldsymbol{M}(s)] \tag{27}$$

On differentiating Eq. (26), we have

$$d\boldsymbol{x}^*(t) = \boldsymbol{A}\int_0^t g'(t-s)[\boldsymbol{m}(s) \odot d\boldsymbol{N}(s) + \boldsymbol{w}(s) \odot d\boldsymbol{M}(s)] + g(0)\boldsymbol{A}[\boldsymbol{m}(t) \odot d\boldsymbol{N}(t) + \boldsymbol{w}(t) \odot d\boldsymbol{M}(t)], \tag{28}$$

where $g(t) = e^{-\omega t}$ and $g'(t-s) = -\omega g(t-s)$, which proves:

$$\begin{aligned} d\mathbf{x}^*(t) &= \omega(\boldsymbol{\alpha} - \mathbf{x}^*(t))dt + \mathbf{A}[\mathbf{m}(t) \odot d\boldsymbol{N}(t) + \boldsymbol{w}(t) \odot d\boldsymbol{M}(t)] \\ \text{Similarly, } d\boldsymbol{\lambda}^*(t) &= \nu(\boldsymbol{\mu} - \boldsymbol{\lambda}^*(t))dt + \mathbf{B}\, d\boldsymbol{N}(t) + \mathbf{B} d\boldsymbol{M}(t) \end{aligned} \tag{29}$$

B Proof of Lemma 2:

For any $u \in \mathcal{O}$, define $\boldsymbol{\theta}_u = [\boldsymbol{A}_u; \alpha_u]$. We observe the loss (Eq. 11) associated with only user u is a regularized least loss, *i.e.*,

$$\hat{\boldsymbol{\theta}}_u = \min_{\boldsymbol{\theta}_u} \sum_{e_i \in \mathcal{H}_T} \sigma^{-2}\Big(m_u^*(t_i) - \boldsymbol{\phi}_i^{uT}\boldsymbol{\theta}_u\Big)^2 + c||\boldsymbol{\theta}_u||_2^2, \tag{30}$$

$$\begin{aligned} \hat{\boldsymbol{\theta}}_u &= \Big(c\boldsymbol{I} + \sigma^{-2}\sum_{e_i \in \mathcal{H}_T} \boldsymbol{\phi}_i^u \boldsymbol{\phi}_i^{uT}\Big)^{-1} \sigma^{-2} \sum_{e_i \in \mathcal{H}_T} m_u(t_i)\boldsymbol{\phi}_i^u \\ &= \Big(c\boldsymbol{I} + \sigma^{-2}\sum_{e_i \in \mathcal{H}_T} \boldsymbol{\phi}_i^u \boldsymbol{\phi}_i^{uT}\Big)^{-1} \sigma^{-2} \sum_{e_i \in \mathcal{H}_T} \boldsymbol{\phi}_i^u \boldsymbol{\phi}_i^{uT}\boldsymbol{\theta}_u + \Big(c\boldsymbol{I} + \sigma^{-2}\sum_{e_i \in \mathcal{H}_T} \boldsymbol{\phi}_i^u \boldsymbol{\phi}_i^{uT}\Big)^{-1} \sum_{e_i \in \mathcal{H}_T} \boldsymbol{\phi}_i^u \frac{\epsilon(t_i)}{\sigma^2} \end{aligned}$$

The equality follows from the fact that, $m_u(t_i) = \boldsymbol{\phi}_i^{uT}\boldsymbol{\theta}_u + \epsilon(t_i)$

$$\hat{\boldsymbol{\theta}}_u - \boldsymbol{\theta}_u = -c\Big(c\boldsymbol{I} + \sigma^{-2}\sum_{e_i \in \mathcal{H}_T} \boldsymbol{\phi}_i^u \boldsymbol{\phi}_i^{uT}\Big)^{-1}\boldsymbol{\theta}_u + \Big(c\boldsymbol{I} + \sigma^{-2}\sum_{e_i \in \mathcal{H}_T} \boldsymbol{\phi}_i^u \boldsymbol{\phi}_i^{uT}\Big)^{-1} \sum_{e_i \in \mathcal{H}_T} \boldsymbol{\phi}_i^u \frac{\epsilon(t_i)}{\sigma^2}$$

Then the covariance product is given in the following:

$$\begin{aligned} \mathbb{E}(\hat{\boldsymbol{\theta}}_u - \boldsymbol{\theta}_u)(\hat{\boldsymbol{\theta}}_u - \boldsymbol{\theta}_u)^T &= c^2\Big(c\boldsymbol{I} + \sigma^{-2}\sum_{e_i \in \mathcal{H}_T} \boldsymbol{\phi}_i^u \boldsymbol{\phi}_i^{uT}\Big)^{-1}\mathbb{E}[\boldsymbol{\theta}_u\boldsymbol{\theta}_u^T]\Big(c\boldsymbol{I} + \sigma^{-2}\sum_{e_i \in \mathcal{H}_T} \boldsymbol{\phi}_i^u \boldsymbol{\phi}_i^{uT}\Big)^{-1} \\ &+ \Big(c\boldsymbol{I} + \sigma^{-2}\sum_{e_i \in \mathcal{H}_T} \boldsymbol{\phi}_i^u \boldsymbol{\phi}_i^{uT}\Big)^{-1} \sum_{e_i \in \mathcal{H}_T} \boldsymbol{\phi}_i^u \boldsymbol{\phi}_i^{uT} \frac{\mathbb{E}(\epsilon^2(t_i))}{\sigma^4}\Big(c\boldsymbol{I} + \sigma^{-2}\sum_{e_i \in \mathcal{H}_T} \boldsymbol{\phi}_i^u \boldsymbol{\phi}_i^{uT}\Big)^{-1} \end{aligned}$$

Note that, from the regularizer we have, $\boldsymbol{\theta}_u \sim \mathcal{N}(0, \boldsymbol{I}/c)$. Furthermore $\mathbb{E}(\epsilon^2(t_i)) = \sigma^2$. Using simple algebraic calculation, we have the value for $\mathbb{\Sigma}(\mathcal{H}_T, OO)$.

C Proof of Theorem 4:

Proof of part (i):

Monotonicity of $f|\mathcal{O}$ **:** To prove monotonicity, we need to show, $f(\mathcal{H}_T \cup \{e_k\}, \mathcal{O}) - f(\mathcal{H}_T) \geq 0$. Assume the user u has posted the event $e_k = \{u_k, m_k, t_k\}$, ***i.e.*** $u_k = u$.

Let the corresponding vector representing the effect of history on e_k on node u is $\boldsymbol{\phi}_k^u$ (as defined in Lemma 2). We define an auxiliary function:

$$g_1(p) := \sum_{u\in\mathcal{O}} \operatorname{tr}\log\Big(c\boldsymbol{I} + \sigma^{-2}\sum_{e_i\in\mathcal{H}_T}\boldsymbol{\phi}_i^u\boldsymbol{\phi}_i^{uT} + p\boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT}\Big), p\in[0,1].$$

Note that, monotonicity of f in $\mathcal{H}_T$ is equivalent to the condition: $g(1) \geq g(0)$ that we would show by proving $\frac{d}{dp}g_1(p) \geq 0$ For compactness, we define

$$\boldsymbol{G}_u = \Big(c\boldsymbol{I} + \sigma^{-2}\sum_{e_i\in\mathcal{H}_T}\boldsymbol{\phi}_i^u\boldsymbol{\phi}_i^{uT}\Big)$$

. Now observe

$$\begin{aligned}
\frac{d}{dp}g_1(p) &= \sum_{u\in\mathcal{O}} \operatorname{tr}\frac{d}{dp}\log(\boldsymbol{G}_u + p\boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT})\\
&= \sum_{u\in\mathcal{O}} \operatorname{tr}\frac{d}{dp}\Big[\log(\boldsymbol{G}_u + p\boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT}) - \log\boldsymbol{G}_u\Big]\\
&= \sum_{u\in\mathcal{O}} \operatorname{tr}\,[(\boldsymbol{I} + p\boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT}\boldsymbol{G}_u^{-1})^{-1}\boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT}\boldsymbol{G}_u^{-1}]\\
&= \sum_{u\in\mathcal{O}} \boldsymbol{\phi}_k^{uT}(\boldsymbol{G}_u + p\boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT})^{-1}\boldsymbol{\phi}_k^u \geq 0 \qquad (31)
\end{aligned}$$

Submodularity of $f|\mathcal{O}$**:** Here we prove submodularity of $f|\mathcal{O}$, whenever $\mathcal{H}_T \subseteq \overline{\mathcal{H}}_T$.

$$f(\mathcal{H}_T\cup e_k, \mathcal{O}) - f(\mathcal{H}_T, \mathcal{O}) \geq f(\overline{\mathcal{H}}_T\cup e_k, \mathcal{O}) - f(\overline{\mathcal{H}}_T, \mathcal{O}) \qquad (32)$$

We assume $f(\overline{\mathcal{H}}_T, \mathcal{O}) = \sum_{u\in\mathcal{O}}\operatorname{tr}\log\overline{\boldsymbol{G}}_u$, where we define $\overline{\boldsymbol{G}}_u$ and also two new matrices: $\boldsymbol{G}_u$, $\boldsymbol{G}_u^\delta$ as follows:

$$\overline{\boldsymbol{G}}_u = \left(\overbrace{c\boldsymbol{I} + \sigma^{-2}\sum_{e_i\in\mathcal{H}_T}\boldsymbol{\phi}_i^u\boldsymbol{\phi}_i^{uT}}^{\boldsymbol{G}_u} + \overbrace{\sigma^{-2}\sum_{e_j\in\mathcal{H}'_T\setminus\mathcal{H}_T}\boldsymbol{\phi}_j^u\boldsymbol{\phi}_j^{uT}}^{\boldsymbol{G}_u^\delta}\right) \qquad (33)$$

That is, $\overline{\boldsymbol{G}}_u = \boldsymbol{G}_u + \boldsymbol{G}_u^\delta$. Now we define:

$$g_2(p) = \operatorname{tr}\log(\boldsymbol{G}_u + p\boldsymbol{G}_u^\delta + \boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT}) - \operatorname{tr}\log(\boldsymbol{G}_u + p\boldsymbol{G}_u^\delta),\ p\in[0,1] \qquad (34)$$

The condition for submodularity (Eq. (32)) is equivalent to $g_2(0) > g_2(1)$. To prove the latter, it is sufficient to prove $\frac{d}{dp}g_2(p) < 0$

$$\begin{aligned}
\frac{d}{dp}g_2(p) &= \operatorname{tr}\frac{d}{dp}\log(\boldsymbol{G}_u + p\boldsymbol{G}_u^\delta + \boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT}) - \operatorname{tr}\frac{d}{dp}\log(\boldsymbol{G}_u + p\boldsymbol{G}_u^\delta)\\
&= \operatorname{tr}\left[\frac{d}{dp}\Big(\log(\boldsymbol{G}_u + p\boldsymbol{G}_u^\delta + \boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT}) - \log(\boldsymbol{G}_u + \boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT})\Big) - \frac{d}{dp}\Big(\log(\boldsymbol{G}_u + p\boldsymbol{G}_u^\delta) - \log\boldsymbol{G}_u\Big)\right]\\
&= \operatorname{tr}\frac{d}{dp}\left[\log\Big(\boldsymbol{I} + p\boldsymbol{G}_u^\delta(\boldsymbol{G}_u + \boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT})^{-1}\Big) - \Big(\log(I + p\boldsymbol{G}_u^\delta\boldsymbol{G}_u^{-1}\Big)\right]\\
&= \operatorname{tr}\left[\Big(\boldsymbol{I} + p\boldsymbol{G}_u^\delta(\boldsymbol{G}_u + \boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT})^{-1}\Big)^{-1}\boldsymbol{G}_u^\delta(\boldsymbol{G}_u + \boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT})^{-1} - \Big(I + p\boldsymbol{G}_u^\delta\boldsymbol{G}_u^{-1}\Big)^{-1}\boldsymbol{G}_u^\delta\boldsymbol{G}_u^{-1}\right]. \qquad (35)
\end{aligned}$$

Using the identity $(\boldsymbol{I}+\boldsymbol{X})^{-1}\boldsymbol{X} = \boldsymbol{I} - (\boldsymbol{I}+\boldsymbol{X})^{-1}$, we have

$$\frac{d}{dp} g_2(p) = \operatorname{tr}\left(I + p\boldsymbol{G}_u^\delta \boldsymbol{G}_u^{-1}\right)^{-1} - \operatorname{tr}\left(\boldsymbol{I} + p\boldsymbol{G}_u^\delta(\boldsymbol{G}_u + \boldsymbol{\phi}_k^u \boldsymbol{\phi}_k^{uT})^{-1}\right)^{-1} \tag{36}$$

Now we use Woodbury matrix identity:

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

to show

$$\begin{aligned} \frac{d}{dp} g_2(p) &= -p \operatorname{tr}\left[\boldsymbol{G}_u^\delta(\boldsymbol{G}_u + p\boldsymbol{G}_u^\delta)^{-1} - \boldsymbol{G}_u^\delta(\boldsymbol{G}_u + p\boldsymbol{G}_u^\delta + \boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT})^{-1}\right] \\ &= -\sigma^{-2} p \sum_{e_j \in \mathcal{H}'_T \setminus \mathcal{H}_T} \boldsymbol{\phi}_j^{uT} \left[(\boldsymbol{G}_u + p\boldsymbol{G}_u^\delta)^{-1} - (\boldsymbol{G}_u + p\boldsymbol{G}_u^\delta + \boldsymbol{\phi}_k^u\boldsymbol{\phi}_k^{uT})^{-1}\right]\boldsymbol{\phi}_j^u \leq 0. \end{aligned} \tag{37}$$

The last inequality is due to $\boldsymbol{G}_u^\delta = \sigma^{-2}\sum_{e_j \in \mathcal{H}'_T \setminus \mathcal{H}_T} \boldsymbol{\phi}_j^u \boldsymbol{\phi}_j^{uT}$.
Now, submodularity of $f|\mathcal{H}_T$ in $\mathcal{O}$ can be easily shown using $f(\mathcal{H}_T, \mathcal{O}) = \sum_{u\in\mathcal{O}} f(\mathcal{H}_T, u)$. And the monotonicity follows from $f(\mathcal{H}_T, \mathcal{O}) > 0$ for $c > 1$.
Proof of part (ii): We re-explain the underlying assumption that, $\mathcal{V}(\mathcal{H}_T) \subseteq \mathcal{O}$; in words, the users of message set $\mathcal{H}_T$ belong to $\mathcal{O}$. Otherwise, we would always be able to find a set $\mathcal{H}'_T \subset \mathcal{H}_T$, so that $\mathcal{V}(\mathcal{H}'_T) \subseteq \mathcal{O}$, with $f(\mathcal{H}_T, \mathcal{O}) = f(\mathcal{H}'_T, \mathcal{O})$. Suppose $v \in \mathcal{V}(\mathcal{H}_T)$ but $v \notin \mathcal{O}$. Since $v \notin \mathcal{O}$, the events posted by user v, *i.e.* the vectors $\boldsymbol{\phi}_i^v$ are not contributing to $f(\mathcal{H}_T, \mathcal{O})$. Therefore $f(\mathcal{H}_T \backslash \{e_v\}, \mathcal{O}) = f(\mathcal{H}_T, \mathcal{O})$, where e_v is an message posted by user v. We consider two cases; (a) u posted e or (b) some user $v \neq u$ posted e. For case (a) $f(\mathcal{H}_T \cup e, \mathcal{O} \cup u) - f(\mathcal{H}_T, \mathcal{O}) = f(e, u)$ which is same as, $f(\overline{\mathcal{H}}_T \cup e, \overline{\mathcal{O}} \cup u) - f(\overline{\mathcal{H}}_T, \overline{\mathcal{O}})$.
For case (b)

$$\begin{aligned} f(\mathcal{H}_T \cup e, \mathcal{O} \cup u) - f(\mathcal{H}_T, \mathcal{O}) &= \sum_{w \neq v} f(\mathcal{H}_T, w) + f(\mathcal{H}_T \cup e, v) - f(\mathcal{H}_T, \mathcal{O}) \\ &= \sum_{w \neq v} f(\mathcal{H}_T, w) + f(\mathcal{H}_T \cup e, v) - \sum_{w \in \mathcal{O}} f(\mathcal{H}_T, w) \\ &= f(\mathcal{H}_T \cup e, v) - f(\mathcal{H}_T, v) \end{aligned} \tag{38}$$

Similarly,

$$f(\overline{\mathcal{H}}_T \cup e, \overline{\mathcal{O}} \cup u) - f(\overline{\mathcal{H}}_T, \overline{\mathcal{O}}) = f(\overline{\mathcal{H}}_T \cup e, v) - f(\mathcal{H}_T, v) \tag{39}$$

The submodularity of $f|v$ suggests that

$$f(\mathcal{H}_T \cup e, v) - f(\mathcal{H}_T, v) \geq f(\overline{\mathcal{H}}_T \cup e, v) - f(\overline{\mathcal{H}}_T, v)$$

which concludes the proof.

C.1 Approximation bound

We restate Lemma 5 here.

Lemma 9 *Algorithm 1 admits an* $(1 - 1/e)$ *approximation bound.*

Suppose $\mathcal{H}_T^* = \{e_k^* | 1 \leq k \leq N_{\mathcal{H}}\}$ and $\mathcal{O}^* = \{v_k^* | 1 \leq k \leq N_{\mathcal{O}}\}$ are the optimum sets. Furthermore, assume $E_i = \{e_1^*, e_2^*, \ldots, e_i^*\}$ and $V_j = \{v_1^*, v_2^*, \ldots, v_j^*\}$ for $1 \leq i \leq N_{\mathcal{H}}$ and $1 \leq j \leq N_{\mathcal{O}}$. For uniformity, assume $E_0, V_0 = \emptyset$. $\mathcal{H}_T[k]$, $\mathcal{O}[k]$ as the selected message and user set at k-th update in step 8 of Algorithm 1.

$$f(\mathcal{H}_T^* \cup \mathcal{H}_T[k], \mathcal{O}^* \cup \mathcal{O}[k]) \quad \leq f(\mathcal{H}_T[k], \mathcal{O}[k]) + \sum_{i=1, j=1}^{N_{\mathcal{H}}, N_{\mathcal{O}}} \Big(f(\mathcal{H}_T[k] \cup E_i, \mathcal{O}[k] \cup V_j) - f(\mathcal{H}_T[k] \cup E_{i-1}, \mathcal{O}[k] \cup V_{j-1})\Big)$$

The second term in the RHS is a telescoping sum which upon subsequent cancellations leads the upper bound above. Now due to joint submodularity, the RHS above is atmost

$$f(\mathcal{H}_T[k], \mathcal{O}[k]) + \sum_{i=1,j=1}^{N_{\mathcal{H}},N_{\mathcal{O}}} \Big(f(\mathcal{H}_T[k] \cup e_i^*, \mathcal{O}[k] \cup v_j^*) - f(\mathcal{H}_T[k], \mathcal{O}[k]) \Big) \tag{40}$$

Again, due to the greedy selection process (in particularly line 6 in Algorithm 1), we have:

$$f(\mathcal{H}_T^* \cup \mathcal{H}_T[k], \mathcal{O}^* \cup \mathcal{O}[k]) \leq f(\mathcal{H}_T[k], \mathcal{O}[k]) + N_{\mathcal{O}} N_{\mathcal{H}} \Big(f(\mathcal{H}_T[k+1], \mathcal{O}[k+1]) - f(\mathcal{H}_T[k], \mathcal{O}[k]) \Big) \tag{41}$$

Due to monotonicity of f, $f(\mathcal{H}_T^*, \mathcal{O}^*) \leq f(\mathcal{H}_T^* \cup \mathcal{H}_T[k], \mathcal{O}^* \cup \mathcal{O}[k])$. So, we have:

$$f(\mathcal{H}_T^*, \mathcal{O}^*) \leq f(\mathcal{H}_T[k], \mathcal{O}[k]) + N_{\mathcal{O}} N_{\mathcal{H}} \Big(f(\mathcal{H}_T[k+1], \mathcal{O}[k+1]) - f(\mathcal{H}_T[k], \mathcal{O}[k]) \Big)$$

Subtracting $N_{\mathcal{O}} N_{\mathcal{H}} f(\mathcal{H}_T, \mathcal{O})$ from both sides, we have:

$$f(\mathcal{H}_T[k+1], \mathcal{O}[k+1]) - f(\mathcal{H}_T^*, \mathcal{O}^*) \geq \frac{N_{\mathcal{O}} N_{\mathcal{H}} - 1}{N_{\mathcal{O}} N_{\mathcal{H}}} \Big(f(\mathcal{H}_T[k], \mathcal{O}[k]) - f(\mathcal{H}_T^*, \mathcal{O}^*) \Big)$$

which implies

$$f(\mathcal{H}_T[k+1], \mathcal{O}[k+1]) \geq \Big(1 - \frac{1}{N_{\mathcal{O}} N_{\mathcal{H}}} \Big)^{k+1} f(\mathcal{H}_T^*, \mathcal{O}^*) \geq (1 - 1/e) f(\mathcal{H}_T^*, \mathcal{O}^*)$$

. The finally selected set $\mathcal{H}_T$ and $\mathcal{O}$ are supersets of $\mathcal{H}_T[k+1]$ and $\mathcal{O}[k+1]$ respectively, which completes the Lemma.

D Proof of Lemma 7

We first restate Lemma 7.
Lemma 7: The HJB equation given by Eq. 20 on expanding satisfies the following:

$$\min_{\boldsymbol{\eta}^{\pm}(t,t+dt]} \Big\{ \frac{\partial J}{\partial t} - \omega(\boldsymbol{x}^*(t) - \boldsymbol{\alpha})^T \frac{\partial J}{\partial \boldsymbol{x}^*} - \nu(\boldsymbol{\lambda}^*(t) - \boldsymbol{\beta})^T \frac{\partial J}{\partial \boldsymbol{\lambda}^*}$$
$$+ \sum_{i\in\mathcal{V}} \lambda_i^*(t) \mathbb{E} \Delta_{\boldsymbol{A}_i m_i, \boldsymbol{B}_i} J + \sum_{i\in\mathcal{I}} \big[\boldsymbol{\eta}_i^+(t) \Delta_{\boldsymbol{A}_i,\boldsymbol{B}_i} J + \boldsymbol{\eta}_i^-(t) \Delta_{-\boldsymbol{A}_i,\boldsymbol{B}_i} J \big]$$
$$+ \ell(\boldsymbol{x}^*(t), \boldsymbol{\eta}^{\pm}(t)) \Big\} = 0 \tag{42}$$

where $\Delta_{\boldsymbol{a},\boldsymbol{b}} J$ is given by $\Delta_{\boldsymbol{a},\boldsymbol{b}} J = J(\boldsymbol{x}^*(t) + \boldsymbol{a}, \boldsymbol{\lambda}^*(t) + \boldsymbol{b}, t) - J(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), t)$ and the expectation is taken over the noise induced due to sampling of m_i from $x_i^*(t)$.
Proof Using the definition of derivative we can evaluate the differential of the cost-to-go as follows:

$$\begin{aligned} dJ(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), t) &= J(\boldsymbol{x}^*(t+dt), \boldsymbol{\lambda}^*(t+dt), t+dt) - J(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), t) \\ &= J(\boldsymbol{x}^*(t) + d\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t) + d\boldsymbol{\lambda}^*(t), t+dt) - J(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), t). \end{aligned}$$

To compute the first term in the RHS of the above equality, we substitute $d\boldsymbol{x}^*(t)$ and $d\boldsymbol{\lambda}^*(t)$ by the SDE dynamics (Eq. 29). Then, using the zero-one jump law [27] we can write:

$$\begin{aligned} &J(\boldsymbol{x}^*(t) + d\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t) + d\boldsymbol{\lambda}^*(t), t+dt) \\ &= J(\boldsymbol{x}^*(t) + f(t)dt + \boldsymbol{A}\, d\boldsymbol{N}(t) + \boldsymbol{C}\,(d\boldsymbol{M}^+(t) - d\boldsymbol{M}^-(t)), \boldsymbol{\lambda}^*(t) + g(t)\,dt + \boldsymbol{A}\, d\boldsymbol{N}(t) + \boldsymbol{D}\,(d\boldsymbol{M}^+(t) + d\boldsymbol{M}^-(t)), t+dt) \\ &= \sum_{i\in\mathcal{V}} J(\boldsymbol{x}^*(t) + f(t)\,dt + \boldsymbol{A}_i, \boldsymbol{\lambda}^*(t) + g(t)dt + \boldsymbol{B}_i, t+dt) dN_i(t) \\ &\quad + \sum_{i\in\mathcal{I}} J(\boldsymbol{x}^*(t) + f(t)\,dt + \boldsymbol{C}_i, \boldsymbol{\lambda}^*(t) + g(t)dt + \boldsymbol{D}_i, t+dt) dM_i^+(t) \\ &\quad + \sum_{i\in\mathcal{I}} J(\boldsymbol{x}^*(t) + f(t)\,dt - \boldsymbol{C}_i, \boldsymbol{\lambda}^*(t) + g(t)dt + \boldsymbol{B}_i, t+dt) dM_i^- \\ &\quad + J(\boldsymbol{x}^*(t) + f(t)dt, \boldsymbol{\lambda}^*(t) + g(t)dt, t+dt) \prod_{i\in\mathcal{V}} [1 - dN_i(t)] \prod_{i\in\mathcal{I}} [1 - dM_i^+(t)] \prod_{i\in\mathcal{I}} [1 - dM_i^-(t)]. \end{aligned}$$

The last expression equals to,

$$J(\boldsymbol{x}^*(t) + f(t)dt, \boldsymbol{\lambda}^*(t) + g(t)dt, t+dt)[1 - \sum_{i\in\mathcal{V}} dN_i(t) - \sum_{i\in\mathcal{I}} (dM_i^+(t) + dM_i^-(t)]$$

Note $f(t) = -\omega(\boldsymbol{x}^*(t) - \boldsymbol{\alpha})$ and $g(t) = -\nu(\boldsymbol{\lambda}^*(t) - \boldsymbol{\mu})$.

$$\begin{aligned} &J(\boldsymbol{x}^*(t) + d\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t) + d\boldsymbol{\lambda}^*(t), t+dt) = J(\boldsymbol{x}^*(t) + f(t)dt, \boldsymbol{\lambda}^*(t) + g(t)dt, t+dt) \\ &\quad + \sum_{i\in\mathcal{V}} \Big[J(\boldsymbol{x}^*(t) + f(t)\,dt + \boldsymbol{A}_i, \boldsymbol{\lambda}^*(t) + g(t)dt + \boldsymbol{B}_i, t+dt) - J(\boldsymbol{x}^*(t) + f(t)dt, \boldsymbol{\lambda}^*(t) + g(t)dt, t+dt) \Big] dN_i(t) \\ &\quad + \sum_{i\in\mathcal{I}} \Big[J(\boldsymbol{x}^*(t) + f(t)\,dt + \boldsymbol{C}_i, \boldsymbol{\lambda}^*(t) + g(t)dt + \boldsymbol{D}_i, t+dt) - J(\boldsymbol{x}^*(t) + f(t)dt, \boldsymbol{\lambda}^*(t) + g(t)dt, t+dt) \Big] dM_i^+(t) \\ &\quad + \sum_{i\in\mathcal{I}} \Big[J(\boldsymbol{x}^*(t) + f(t)\,dt - \boldsymbol{C}_i, \boldsymbol{\lambda}^*(t) + g(t)dt + \boldsymbol{B}_i, t+dt) - J(\boldsymbol{x}^*(t) + f(t)dt, \boldsymbol{\lambda}^*(t) + g(t)dt, t+dt) \Big] dM_i^-(t) \end{aligned}$$

We use the bilinear differential form $dt\, dN(t) = 0$ as in [23] to reduce $-J(\boldsymbol{x}^*(t) + f(t)dt, \boldsymbol{\lambda}^*(t) + g(t)dt, t+dt)$ in all the last three terms that are driven by jumps, into $J(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), t)$. Furthermore, we apply chain rule on the first term $J(\boldsymbol{x}^*(t)+f(t)dt, \boldsymbol{\lambda}^*(t)+g(t)dt, t+dt)$ to have, the following

$$J(\boldsymbol{x}^*(t)+d\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t)+d\boldsymbol{\lambda}^*(t), t+dt) = J(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), t) + \left[\frac{\partial J}{\partial t} + f^T(t)\frac{\partial J}{\partial \boldsymbol{x}^*} + g^T(t)\frac{\partial J}{\partial \boldsymbol{\lambda}^*}\right] dt$$
$$+ \sum_{i\in\mathcal{V}} dN_i(t)\Delta_{\boldsymbol{A}_i m_i, \boldsymbol{B}_i} J + \sum_{i\in\mathcal{I}} \left[dM_i^+(t)\Delta_{\boldsymbol{A}_i,\boldsymbol{B}_i} J + dM_i^-(t)\Delta_{-\boldsymbol{A}_i,\boldsymbol{B}_i} J\right]$$

If we plug the values of $f(t)$ and $g(t)$ into the previous one, take expectation with repect to the $\boldsymbol{N}(t)$ and $\boldsymbol{M}(t)$, and finally put them to Eq. 20to complete the proof. ■

E Proof of Lemma 8

We express the required terms as following:

$$\frac{\partial J}{\partial t} = \dot{h}(t) + \dot{\boldsymbol{g}}(t)^T\boldsymbol{\lambda}^*(t) + \dot{\boldsymbol{f}}^T(t)\boldsymbol{x}^*(t) + \boldsymbol{\lambda}^{*T}(t)\dot{\boldsymbol{V}}(t)\boldsymbol{x}^*(t) + \frac{1}{2}\boldsymbol{\lambda}^{*T}(t)\dot{\boldsymbol{K}}(t)\boldsymbol{\lambda}^*(t) + \frac{1}{2}\boldsymbol{x}^{*T}(t)\dot{\boldsymbol{H}}(t)\boldsymbol{x}^*(t) \tag{43}$$

$$-\omega(\boldsymbol{x}^*(t)-\boldsymbol{\alpha})^T\frac{\partial J}{\partial \boldsymbol{x}^*} = -\omega\boldsymbol{f}^T(t)(\boldsymbol{x}^*(t)-\boldsymbol{\alpha}) - \omega\boldsymbol{\lambda}^{*T}(t)\boldsymbol{V}(t)(\boldsymbol{x}^*(t)-\boldsymbol{\alpha}) - \omega(\boldsymbol{x}^*(t)-\boldsymbol{\alpha})^T\boldsymbol{H}(t)\boldsymbol{x}^*(t) \tag{44}$$

$$-\nu(\boldsymbol{\lambda}^*(t)-\boldsymbol{\mu})^T\frac{\partial J}{\partial \boldsymbol{\lambda}^*} = -\nu\boldsymbol{g}^T(t)(\boldsymbol{\lambda}^*(t)-\boldsymbol{\mu}) - \nu(\boldsymbol{\lambda}^*(t)-\boldsymbol{\mu})^T\boldsymbol{V}(t)\boldsymbol{x}^{*T}(t) - \nu(\boldsymbol{\lambda}^*(t)-\boldsymbol{\mu})^T\boldsymbol{K}(t)\boldsymbol{\lambda}^*(t) \tag{45}$$

$$\begin{aligned}\sum_{i\in\mathcal{V}}\lambda_i^*(t)\mathbb{E}\Delta_{\boldsymbol{A}_i m_i,\boldsymbol{B}_i}J = \sum_{i\in\mathcal{V}}\lambda_i(t)\mathbb{E}_{m_i|x_i}\Big[&m_i\boldsymbol{A}_i^T\boldsymbol{H}(t)\boldsymbol{x}^*(t) + \frac{1}{2}m_i^2\boldsymbol{A}_i^T\boldsymbol{H}(t)\boldsymbol{A}_i + \boldsymbol{B}_i^T\boldsymbol{K}(t)\boldsymbol{\lambda}^*(t) + \frac{1}{2}\boldsymbol{B}_i^T\boldsymbol{K}(t)\boldsymbol{B}_i\\ &+ m_i\boldsymbol{\lambda}^{*T}(t)\boldsymbol{V}(t)\boldsymbol{A}_i + \boldsymbol{B}_i^T\boldsymbol{V}(t)\boldsymbol{x}(t) + m_i\boldsymbol{B}_i^T\boldsymbol{V}(t)\boldsymbol{A}_i + \boldsymbol{g}^T(t)\boldsymbol{B}_i + m_i\boldsymbol{f}^T(t)\boldsymbol{A}_i\Big]\end{aligned} \tag{46}$$

$$\begin{aligned}\Delta_{\pm\boldsymbol{A}_i,\boldsymbol{B}_i}J = &\pm \boldsymbol{x}^{*T}(t)\boldsymbol{H}(t)\boldsymbol{A}_i + \frac{1}{2}\boldsymbol{A}_i^T\boldsymbol{H}(t)\boldsymbol{A}_i + \boldsymbol{\lambda}^{*T}(t)\boldsymbol{K}(t)\boldsymbol{B}_i + \frac{1}{2}\boldsymbol{B}_i^T\boldsymbol{K}(t)\boldsymbol{B}_i\\ &\pm \boldsymbol{\lambda}^{*T}(t)\boldsymbol{V}(t)\boldsymbol{A}_i + \boldsymbol{B}_i^T\boldsymbol{V}(t)\boldsymbol{x}^*(t) + \boldsymbol{B}_i^T\boldsymbol{V}(t)\boldsymbol{A}_i + \boldsymbol{g}^T(t)\boldsymbol{B}_i \pm \boldsymbol{f}^T(t)\boldsymbol{A}_i\end{aligned} \tag{47}$$

From the last equation Eq. 47

$$\begin{aligned}\frac{1}{2}\sum_{s\in\pm}\Delta_{s\boldsymbol{A},\boldsymbol{B}}J^T\boldsymbol{S}^{-1}\Delta_{s\boldsymbol{A},\boldsymbol{B}}J = -\sum_{\substack{i\in\mathcal{I}\\ s\in\pm}}\frac{1}{2s_i}\Big[&s\boldsymbol{x}^{*T}(t)\boldsymbol{H}(t)\boldsymbol{A}_i + \frac{1}{2}\boldsymbol{A}_i^T\boldsymbol{H}(t)\boldsymbol{A}_i + \boldsymbol{\lambda}^{*T}(t)\boldsymbol{K}(t)\boldsymbol{B}_i + \frac{1}{2}\boldsymbol{B}_i^T\boldsymbol{K}(t)\boldsymbol{B}_i\\ &+ s\boldsymbol{\lambda}^{*T}(t)\boldsymbol{V}(t)\boldsymbol{A}_i + \boldsymbol{B}_i^T\boldsymbol{V}(t)\boldsymbol{x}^*(t) + \boldsymbol{B}_i^T\boldsymbol{V}(t)\boldsymbol{A}_i + \boldsymbol{g}^T(t)\boldsymbol{B}_i + s\boldsymbol{f}^T(t)\boldsymbol{A}_i\Big]^2\end{aligned}$$

which is same as,

$$\begin{aligned}-\sum_{i\in\mathcal{I}}\frac{1}{s_i}\Big[&\boldsymbol{x}^{*T}(t)\Big(\boldsymbol{H}(t)\boldsymbol{A}_i\boldsymbol{A}_i^T\boldsymbol{H}(t) + \boldsymbol{V}(t)\boldsymbol{B}_i\boldsymbol{B}_i^T\boldsymbol{V}(t)\Big)\boldsymbol{x}^*(t) + \boldsymbol{\lambda}^{*T}(t)\Big(\boldsymbol{V}(t)\boldsymbol{A}_i\boldsymbol{A}_i^T\boldsymbol{V}(t) + \boldsymbol{K}(t)\boldsymbol{B}_i\boldsymbol{B}_i^T\boldsymbol{K}(t)\Big)\boldsymbol{\lambda}^*(t)\\ &+ 2\boldsymbol{\lambda}^{*T}(t)\Big(\boldsymbol{V}(t)\boldsymbol{A}_i\boldsymbol{A}_i^T\boldsymbol{H}(t) + \boldsymbol{Q}\boldsymbol{B}_i\boldsymbol{B}_i^T\boldsymbol{V}(t)\Big)\boldsymbol{x}^*(t) + 2\Big(k_i\boldsymbol{A}_i^T\boldsymbol{H}(t) + z_i\boldsymbol{B}_i^T\boldsymbol{V}(t)\Big)\boldsymbol{x}^*(t)\\ &+2\Big(k_i\boldsymbol{A}_i^T\boldsymbol{V}(t) + z_i\boldsymbol{B}_i^T\boldsymbol{K}(t)\Big)\boldsymbol{\lambda}^*(t) + k_i^2 + z_i^2\Big]\end{aligned}$$

Here $k_i = \boldsymbol{B}_i^T \boldsymbol{V}(t)\boldsymbol{A}_i + \boldsymbol{f}^T(t)\boldsymbol{A}_i$ and $z_i = \frac{1}{2}\boldsymbol{A}_i^T \boldsymbol{H}(t)\boldsymbol{A}_i + \frac{1}{2}\boldsymbol{B}_i^T \boldsymbol{K}(t)\boldsymbol{B}_i + \boldsymbol{B}_i^T \boldsymbol{V}(t)\boldsymbol{A}_i$. Now we put Eq. (43) to (47) in Eq. (22), and then equate the coefficients of different terms.

Equating coefficients of quadratics in $\boldsymbol{x}^*(t)$

$$\dot{\boldsymbol{H}}(t) - (\omega \boldsymbol{I} - \boldsymbol{A}\boldsymbol{\Lambda})^T \boldsymbol{H}(t) - \boldsymbol{H}(t)(\omega \boldsymbol{I} - \boldsymbol{A}\boldsymbol{\Lambda}) + \mathrm{diag}(\boldsymbol{A}^T \boldsymbol{H}(t)\boldsymbol{A}) \odot \mathrm{diag}(\boldsymbol{\Lambda})$$
$$- \sum_{i\in\mathcal{I}} \frac{1}{s_i}(\boldsymbol{H}(t)\boldsymbol{A}_i\boldsymbol{A}_i^T\boldsymbol{H}(t) + \boldsymbol{V}(t)\boldsymbol{B}_i\boldsymbol{B}_i^T\boldsymbol{V}(t)) + \frac{1}{2}\boldsymbol{Q}_{\text{From loss } \ell(\boldsymbol{x}^*(t),\, \boldsymbol{\eta}^{\pm}(t))} = 0 \tag{48}$$

Here $\boldsymbol{\Lambda} := \mathrm{diag}(\boldsymbol{\lambda}(t))$, the diagonal matrix derived using the vector $\boldsymbol{\lambda}(t)$.

Equating coefficients of quadratics in $\boldsymbol{\lambda}^*(t)$

$$\dot{\boldsymbol{Q}}(t) - (\nu \boldsymbol{I} - \boldsymbol{B})^T \boldsymbol{Q}(t) - \boldsymbol{K}(t)(\nu \boldsymbol{I} - \boldsymbol{B}) - \sum_{i\in\mathcal{I}} \frac{1}{s_i}\Big(\boldsymbol{V}(t)\boldsymbol{A}_i\boldsymbol{A}_i^T\boldsymbol{V}(t) + \boldsymbol{K}(t)\boldsymbol{B}_i\boldsymbol{B}_i^T\boldsymbol{K}(t)\Big) = 0 \tag{49}$$

Equating coefficients of quadratics in $\boldsymbol{\lambda}^*(t),\ \boldsymbol{x}^*(t)$

$$\dot{\boldsymbol{V}}(t) - \boldsymbol{V}(t)(\omega \boldsymbol{I} + \boldsymbol{A}\boldsymbol{\Lambda}) - (\nu \boldsymbol{I} + \boldsymbol{B}^T)\boldsymbol{V}(t) + \mathrm{diag}(\boldsymbol{f}^T(t)\boldsymbol{A}) - \sum_{i\in\mathcal{I}} \frac{1}{s_i}\Big(\boldsymbol{V}(t)\boldsymbol{A}_i\boldsymbol{A}_i^T\boldsymbol{H}(t) + \boldsymbol{K}(t)\boldsymbol{B}_i\boldsymbol{B}_i^T\boldsymbol{V}(t)\Big) \tag{50}$$

Equating coefficients of linear terms in $\boldsymbol{x}^*(t)$

$$\dot{\boldsymbol{f}}^T(t) + \omega\boldsymbol{\alpha}^T\boldsymbol{H}(t) + \nu\boldsymbol{\mu}^T\boldsymbol{V}(t) - \omega\boldsymbol{f}^T(t) - 2\sum_{i\in\mathcal{I}} \frac{1}{s_i}\Big(k_i\boldsymbol{A}_i^T\boldsymbol{H}(t) + z_i\boldsymbol{B}_i^T\boldsymbol{V}(t)\Big) - \boldsymbol{x}^{\text{Track } T}\boldsymbol{Q}_{\text{From } \ell(\boldsymbol{x}^*(t),\, \boldsymbol{\eta}^{\pm}(t))} = 0 \tag{51}$$

Equating coefficients of linear terms in $\boldsymbol{\lambda}^*(t)$

$$\dot{\boldsymbol{g}}^T(t) + \omega\boldsymbol{\alpha}^T\boldsymbol{V}^T(t) + \nu\boldsymbol{\mu}^T\boldsymbol{K}(t) - \nu\boldsymbol{g}^T(t) + \frac{1}{2}\overline{\mathrm{diag}}\Big(\sigma^2\boldsymbol{A}\boldsymbol{H}(t)\boldsymbol{A} + \boldsymbol{B}^T\boldsymbol{K}(t)\boldsymbol{B}\Big) + \boldsymbol{g}^T(t)\boldsymbol{B} + 2\Big(k_i\boldsymbol{A}_i^T\boldsymbol{V}(t) + z_i\boldsymbol{B}_i^T\boldsymbol{K}(t)\Big) \tag{52}$$

Here $\overline{\mathrm{diag}}(X)$ represents the row vector with diagonal elements of X.

Equating constant terms

$$\boldsymbol{h}(t)^T + \omega\boldsymbol{f}^T(t)\boldsymbol{\alpha} + \nu\boldsymbol{g}^T(t)\boldsymbol{\mu} + k_i^2 + z_i^2 + \frac{1}{2}\boldsymbol{x}^{\text{Track } T}\boldsymbol{Q}_{\text{From } \ell(\boldsymbol{x}^*(t),\, \boldsymbol{\eta}^{\pm}(t))}\boldsymbol{x}^{\text{Track}} = 0 \tag{53}$$

Equations 48 to 53 shows the structure of Riccati equation:

$$\{\dot{\boldsymbol{H}}(t), \dot{\boldsymbol{K}}(t), \dot{\boldsymbol{V}}(t), \dot{\boldsymbol{f}}(t), \dot{\boldsymbol{g}}(t), \dot{\boldsymbol{h}}(t)\} = \mathbf{Riccati}(\boldsymbol{H}(t), \boldsymbol{K}(t), \boldsymbol{V}(t), \boldsymbol{f}(t), \boldsymbol{g}(t)) \tag{54}$$