

## A Proof of Proposition 1

Given  $\mathbf{x}^*(t) := [\mathbf{x}_1(t), \dots, \mathbf{x}_{|\mathcal{V}|}(t)]^T$  and  $\boldsymbol{\lambda}^*(t) := [\boldsymbol{\lambda}_1(t), \dots, \boldsymbol{\lambda}_{|\mathcal{V}|}(t)]^T$ , we can compactly rewrite Eqs. 3 and 4 for all users as:

$$\mathbf{x}^*(t) = \boldsymbol{\alpha} + \mathbf{A} \int_0^t g(t - \theta) \mathbf{m}(t) \odot d\mathbf{N}(t) \quad (15)$$

and

$$\boldsymbol{\lambda}^*(t) = \boldsymbol{\mu} + \mathbf{B} \int_0^t \kappa(t - \theta) \odot d\mathbf{N}(t). \quad (16)$$

Then, it easily follows that

$$d\mathbf{x}^*(t) = \mathbf{A} \int_0^t g'(t - \theta) \mathbf{m}(t) \odot d\mathbf{N}(t) + g(0) \mathbf{A} \mathbf{m}(t) \odot d\mathbf{N}(t), \quad (17)$$

where  $g(t) = e^{-\omega t}$  and  $g'(t - \theta) = -\omega g(t - \theta)$ . And, we can rewrite the above expression as

$$d\mathbf{x}^*(t) = \omega(\boldsymbol{\alpha} - \mathbf{x}^*(t))dt + \mathbf{A}(\mathbf{m}(t) \odot d\mathbf{N}(t)). \quad (18)$$

Similarly, we can show that

$$d\boldsymbol{\lambda}^*(t) = \nu(\boldsymbol{\mu} - \boldsymbol{\lambda}^*(t))dt + \mathbf{B} d\mathbf{N}(t).$$

## B Auxiliary theoretical results

The proofs of Theorems 2 and 5 rely on the following auxiliary Lemma.

**Lemma 7** *The expected opinion  $\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t) | \mathcal{H}_{t_0}]$  in the model of opinion dynamics defined by Eqs. 4 and 3 with exponential triggering kernels with parameters  $\omega$  and  $\nu$  satisfies the following differential equation:*

$$\frac{d\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t) | \mathcal{H}_{t_0}]}{dt} = \mathbf{A} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t) \odot \boldsymbol{\lambda}^*(t) | \mathcal{H}_{t_0}] - \omega \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t) | \mathcal{H}_{t_0}] + \omega \boldsymbol{\alpha}, \quad (19)$$

where  $\mathbf{A} = (a_{vu})_{v,u \in \mathcal{G}}$  and the sign  $\odot$  denotes pointwise product.

Using that  $\mathbb{E}[m_v(\theta) | x_v^*(\theta)] = x_v^*(\theta)$ , we can compute the average opinion of user  $u$  across all possible histories from Eq. 4 as

$$\begin{aligned} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[x_u^*(t) | \mathcal{H}_{t_0}] &= \alpha_u + \sum_{v \in \mathcal{N}(u)} a_{uv} \int_0^t g(t - \theta) \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[m_v(\theta) dN_v(\theta) | \mathcal{H}_{t_0}] \\ &= \alpha_u + \sum_{v \in \mathcal{N}(u)} a_{uv} \sum_{t_i \in \mathcal{H}_v(t_0)} g(t - t_i) m_v(t_i) + \sum_{v \in \mathcal{N}(u)} a_{uv} \int_{t_0}^t g(t - \theta) \mathbb{E}_{\mathcal{H}(\theta) \setminus \mathcal{H}_{t_0}}[x_v^*(\theta) \lambda_v^*(\theta) | \mathcal{H}_{t_0}] d\theta, \end{aligned}$$

and we can write the expectation of the opinion for all users in vectorial form as

$$\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)] = \mathbf{v}(t) + \mathbf{A} \int_0^t g(t - \theta) \mathbb{E}_{\mathcal{H}(\theta) \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(\theta) \odot \boldsymbol{\lambda}^*(\theta)] d\theta, \quad (20)$$

where the  $\odot$  denotes pointwise product and

$$(\mathbf{v}(t))_u = \alpha_u + \sum_{v \in \mathcal{N}(u)} a_{uv} \sum_{t_i \in \mathcal{H}_v(t_0)} g(t - t_i) m_v(t_i).$$

Since  $g(t) = e^{-\omega t}$ , one may observe that  $\omega \mathbf{v}(t) + \dot{\mathbf{v}}(t) = \omega \boldsymbol{\alpha}$ . Then, by differentiating Eq. 20, we obtain

$$\frac{d\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t) | \mathcal{H}_{t_0}]}{dt} = \mathbf{A} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t) \odot \boldsymbol{\lambda}^*(t) | \mathcal{H}_{t_0}] - \omega \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t) | \mathcal{H}_{t_0}] + \omega \boldsymbol{\alpha}, \quad (21)$$

## C Proof of Theorem 2

Using Lemma 7 and  $\lambda_u^*(t) = \mu_u$ , we obtain

$$\frac{d\mathbb{E}_{\mathcal{H}_t}[\mathbf{x}^*(t)]}{dt} = [-\omega I + \mathbf{A}\mathbf{\Lambda}_1]\mathbb{E}_{\mathcal{H}_t}[\mathbf{x}^*(t)] + \omega\boldsymbol{\alpha}, \quad (22)$$

where  $\mathbf{\Lambda}_1 = \text{diag}[\boldsymbol{\mu}]$ . Then, we apply the Laplace transform to the expression above and obtain

$$\hat{\mathbf{x}}(s) = [sI + \omega I - \mathbf{A}\mathbf{\Lambda}_1]^{-1}\mathbf{x}(t_0) + \frac{\omega}{s}[sI + \omega I - \mathbf{A}\mathbf{\Lambda}_1]^{-1}\boldsymbol{\alpha},$$

where we leverage the fact that, conditioning the prior history, the opinion is non-random, *i.e.*,  $\mathbb{E}_{\mathcal{H}_{t_0} \setminus \mathcal{H}_{t_0}^-}[\mathbf{x}(t_0)|\mathcal{H}_{t_0}^-] = \mathbf{x}(t_0)$ . Finally, applying the inverse Laplace transform, we obtain the average opinion  $\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)|\mathcal{H}_{t_0}]$  in time domain as

$$\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)|\mathcal{H}_{t_0}] = e^{(\mathbf{A}\mathbf{\Lambda}_1 - \omega I)(t-t_0)}\mathbf{x}(t_0) + \omega(\mathbf{A}\mathbf{\Lambda}_1 - \omega I)^{-1} \left( e^{(\mathbf{A}\mathbf{\Lambda}_1 - \omega I)(t-t_0)} - \mathbf{I} \right) \boldsymbol{\alpha}.$$

## D Proof of Theorem 3

Theorem 2 states that the average users' opinion  $\mathbb{E}_{\mathcal{H}_t}[\mathbf{x}^*(t)]$  in time domain is given by

$$\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)|\mathcal{H}_{t_0}] = e^{(\mathbf{A}\mathbf{\Lambda}_1 - \omega I)(t-t_0)}\mathbf{x}(t_0) + \omega(\mathbf{A}\mathbf{\Lambda}_1 - \omega I)^{-1} \left( e^{(\mathbf{A}\mathbf{\Lambda}_1 - \omega I)(t-t_0)} - \mathbf{I} \right) \boldsymbol{\alpha}.$$

If  $\text{Re}[\xi(\mathbf{A}\mathbf{\Lambda}_1)] < \omega$ , where  $\xi(\mathbf{X})$  denote the eigenvalues of matrix  $\mathbf{X}$ , it easily follows that

$$\lim_{t \rightarrow \infty} \mathbb{E}_{\mathcal{H}_t}[\mathbf{x}^*(t)] = \left( \mathbf{I} - \frac{\mathbf{A}\mathbf{\Lambda}_1}{\omega} \right)^{-1} \boldsymbol{\alpha}. \quad (23)$$

## E Proof of Theorem 4

Assume  $b_{vu} = 0$  for all  $v, u \in \mathcal{G}, v \neq u$ . Then,  $\lambda_v^*(t)$  only depends on user  $v$ 's history and, since  $x_v^*(t)$  only depends on the history of the user  $v$ 's neighbors  $\mathcal{N}(v)$ , we can write

$$\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t) \odot \boldsymbol{\lambda}^*(t)|\mathcal{H}_{t_0}] = \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)|\mathcal{H}_{t_0}] \odot \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\boldsymbol{\lambda}^*(t)],$$

and rewrite Eq. 21 as

$$\begin{aligned} \frac{d\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)|\mathcal{H}_{t_0}]}{dt} = \\ \mathbf{A}(\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)|\mathcal{H}_{t_0}] \odot \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\boldsymbol{\lambda}^*(t)|\mathcal{H}_{t_0}]) - \omega\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)|\mathcal{H}_{t_0}] + \omega\boldsymbol{\alpha}. \end{aligned} \quad (24)$$

We can now compute  $\mathbb{E}_{\mathcal{H}(\theta) \setminus \mathcal{H}_{t_0}}[\boldsymbol{\lambda}^*(\theta)|\mathcal{H}_{t_0}]$  analytically as follows. From Eq. 3, we obtain

$$\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\boldsymbol{\lambda}^*(t)|\mathcal{H}_{t_0}] = \boldsymbol{\eta}(t) + \int_{t_0}^t \mathbf{B}\kappa(t-\theta)\mathbb{E}_{\mathcal{H}_\theta \setminus \mathcal{H}_{t_0}}[\boldsymbol{\lambda}^*(\theta)]d\theta, \quad (25)$$

where  $\kappa(t) = e^{-\nu t}$ ,  $[\boldsymbol{\eta}(t)]_{u \in \mathcal{V}} = \mu_u + \sum_{v \in \mathcal{N}(u)} b_{uv} \sum_{t_i \in \mathcal{H}_v(t_0)} \kappa(t-t_i)$  and  $\mathbf{B} = (b_{vu})_{v, u \in \mathcal{V}}$ , where  $b_{vu} = 0$  for all  $v \neq u$ , by assumption. Differentiating with respect to  $t$ , we get,

$$\frac{d}{dt}\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\boldsymbol{\lambda}^*(t)|\mathcal{H}_{t_0}] = -\nu\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\boldsymbol{\lambda}^*(t)|\mathcal{H}_{t_0}] + \mathbf{B}\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\boldsymbol{\lambda}^*(t)|\mathcal{H}_{t_0}] + \nu\boldsymbol{\mu},$$

with initial condition  $\mathbb{E}_{\mathcal{H}(t_0^+) \setminus \mathcal{H}(t_0)}[\boldsymbol{\lambda}^*(t_0)] = \boldsymbol{\eta}(t_0)$ . By taking the Laplace transform and then applying inverse Laplace transform,

$$\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\boldsymbol{\lambda}^*(t)|\mathcal{H}_{t_0}] = e^{(\mathbf{B} - \nu I)(t-t_0)}\boldsymbol{\eta}(t_0) + \nu(\mathbf{B} - \nu I)^{-1} \left( e^{(\mathbf{B} - \nu I)(t-t_0)} - \mathbf{I} \right) \boldsymbol{\mu} \quad \forall t \geq t_0, \quad (26)$$

where  $\boldsymbol{\eta}(t_0) = \mathbb{E}_{\mathcal{H}(t_0^+) \setminus \mathcal{H}(t_0)}[\boldsymbol{\lambda}^*(t_0)]$ . Using Eqs. 24 and 26, as well as  $\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)] \odot \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\boldsymbol{\lambda}^*(t)] = \mathbf{\Lambda}(t)\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)]$ , where  $\mathbf{\Lambda}(t) := \text{diag}[\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\boldsymbol{\lambda}^*(t)]]$ , we obtain

$$\frac{d\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)]}{dt} = [-\omega I + \mathbf{A}\mathbf{\Lambda}(t)]\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)] + \omega\boldsymbol{\alpha},$$

with initial conditions  $(\mathbb{E}_{\mathcal{H}(t_0^+) \setminus \mathcal{H}(t_0)}[\mathbf{x}^*(t_0)])_{u \in \mathcal{V}} = \alpha_u + \sum_{v \in \mathcal{N}(u)} a_{uv} \sum_{t_i \in \mathcal{H}_v(t_0)} g(t_0 - t_i)m_v(t_i)$ .

## F Proof of Theorem 5

Theorem 4 states that the average users' opinion  $\mathbb{E}_{\mathcal{H}_t}[\mathbf{x}^*(t)]$  in time domain is given by

$$\frac{d\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)|\mathcal{H}_{t_0}]}{dt} = [-\omega I + \mathbf{A}\mathbf{\Lambda}(t)]\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)|\mathcal{H}_{t_0}] + \omega\boldsymbol{\alpha}. \quad (27)$$

In such systems, solutions can be written as [16]

$$\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)|\mathcal{H}_{t_0}] = \Phi(t)\boldsymbol{\alpha} + \omega \int_0^t \Phi(s)\boldsymbol{\alpha} ds, \quad (28)$$

where the transition matrix  $\Phi(t)$  defines as a solution of the matrix differential equation

$$\dot{\Phi}(t) = [-\omega I + \mathbf{A}\mathbf{\Lambda}(t)]\Phi(t) \text{ with } \Phi(0) = I.$$

If  $\Phi(t)$  satisfies  $\|\Phi(t)\| \leq \gamma e^{-ct} \forall t > 0$  for  $\gamma, c > 0$  then the steady state solution to Eq. 28 is given by [16]

$$\lim_{t \rightarrow \infty} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)|\mathcal{H}_{t_0}] = \left(I - \frac{\mathbf{A}\mathbf{\Lambda}_2}{\omega}\right)^{-1} \boldsymbol{\alpha}.$$

where  $\mathbf{\Lambda}_2 = \lim_{t \rightarrow \infty} \mathbf{\Lambda}(t) = \text{diag}\left[I - \frac{\mathbf{B}}{\nu}\right]^{-1} \boldsymbol{\mu}$ .

## G Proof of Theorem 6

Let  $\{\mathbf{x}_l^*(t)\}_{l=1}^n$  be the simulated opinions for all users and define  $\mathbf{s}(t) = \frac{1}{n} \sum_{l=1}^n \mathbf{s}_l(t)$ , where  $\mathbf{s}_l(t) = (\hat{\mathbf{x}}_l^*(t) - \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)|\mathcal{H}_{t_0}])$ . Clearly, for a given  $t$ , all elements in  $\mathbf{s}_l(t)$  are i.i.d. random variables with zero mean and variance, and we can bound  $|s_u(t)| < 2x_{\max}$ . Then, by Bernstein's inequality, the following holds true,

$$\mathbb{P}(|s_u(t)| > \epsilon) = \mathbb{P}(s_u(t) > \epsilon) + \mathbb{P}(s_u(t) < -\epsilon) > 2 \cdot \exp\left(-\frac{3n\epsilon^2}{6\sigma_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}^2(x_u^*(t)|\mathcal{H}_{t_0}) + 4x_{\max}\epsilon}\right)$$

Let  $\sigma_{\max}^2(t) = \max_{u \in \mathcal{G}} \sigma_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}^2(x_u^*(t)|\mathcal{H}_{t_0})$ . If we choose,

$$\delta < 2 \cdot \exp\left(-\frac{3n\epsilon^2}{6\sigma_{\max}^2(t) + 4x_{\max}\epsilon}\right) \quad (29)$$

we obtain the required bound for  $n$ . Moreover, given this choice of  $\delta$ , we have  $\mathbb{P}(|s_u(t)| < \epsilon) > 1 - \delta$  immediately.

However, a finite bound on  $n$  requires the variance  $\sigma_{\max}^2(t)$  to be bounded for all  $t$ . Hence, we analyze the variance and its stability below.

### G.1 Dynamics of variance

In this section we compute the time-domain evolution of the variance and characterize its stability for Poisson driven opinion dynamics. A general analysis of the variance for multidimensional Hawkes is left for future work.

**Lemma 8** *Given a collection of messages  $\mathcal{H}_{t_0}$  recorded during a time period  $[0, t_0)$  and  $\lambda_u^*(t) = \mu_u$  for all  $u \in \mathcal{G}$ , the covariance matrix  $\mathbf{\Gamma}(t_0, t)$  at any time  $t$  conditioned on the history  $\mathcal{H}_{t_0}$  can be described as,*

$$\text{vec}(\mathbf{\Gamma}(t_0, t)) = \int_0^t \boldsymbol{\Phi}(t - \theta) \text{vec}[\sigma^2 \mathbf{A}\mathbf{\Lambda}\mathbf{A}^T + \mathbf{A} \text{diag}(\mathbb{E}_{\mathcal{H}_{t-}}[\mathbf{x}^*(\theta)]^2 \mathbf{\Lambda}\mathbf{A}^T) d\theta.$$

where

$$\boldsymbol{\Phi}(t) = e^{((-\omega I + \mathbf{A}\mathbf{\Lambda}) \otimes I + I \otimes (-\omega I + \mathbf{A}\mathbf{\Lambda}) + (\mathbf{A} \otimes \mathbf{A}) \hat{\mathbf{\Lambda}})t},$$

$\hat{\mathbf{\Lambda}}_{i^2, i^2} = \boldsymbol{\lambda}^*(i)$ , and  $\mathbf{\Lambda} := \text{diag}[\boldsymbol{\lambda}]$ . Moreover, the stability of the system is characterized by

$$\xi[(-\omega I + \mathbf{A}\mathbf{\Lambda}) \otimes I + I \otimes (-\omega I + \mathbf{A}\mathbf{\Lambda}) + (\mathbf{A} \otimes \mathbf{A}) \hat{\mathbf{\Lambda}}] < 0.$$

**Proof.** By definition, the covariance matrix is given by

$$\mathbf{\Gamma}(t_0, t) := \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}} [(\mathbf{x}^*(t) - \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t)))(\mathbf{x}^*(t) - \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t)))^T | \mathcal{H}_{t_0}]. \quad (30)$$

Hence, if we define  $\Delta \mathbf{x} = (\mathbf{x}^*(t) - \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t)))$ , we can compute the differential of the covariance matrix as

$$d\mathbf{\Gamma}(t_0, t) = \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}} [d(\Delta \mathbf{x} \Delta \mathbf{x}^T) | \mathcal{H}_{t_0}] = \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}} [\Delta \mathbf{x} d(\Delta \mathbf{x}^T) + d(\Delta \mathbf{x}) \Delta \mathbf{x}^T + d(\Delta \mathbf{x}) d(\Delta \mathbf{x}^T) | \mathcal{H}_{t_0}], \quad (31)$$

where

$$d(\Delta \mathbf{x}) = d(\mathbf{x}^*(t) - \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t))) = d(\mathbf{x}^*(t)) - d(\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t))). \quad (32)$$

Next, note that

$$\mathbb{E}[d\mathbf{N}(t)d\mathbf{N}^T(t)] = \mathbb{E}[dN_i(t)dN_j(t)]_{i,j \in \mathcal{V}} = \mathbb{E}[\text{diag}(d\mathbf{N}(t))] = \mathbf{\Lambda}, \quad (33)$$

where the off-diagonal entries vanish, since two jumps cannot happen at the same time point [14].

Now, recall the Markov representation of our model, *i.e.*,

$$d\mathbf{x}^*(t) = -\omega \mathbf{x}^*(t)dt + \mathbf{A}\mathbf{M}^*(t)d\mathbf{N}(t) + \boldsymbol{\alpha}dt, \quad (34)$$

where  $\mathbf{M}^*(t) := \text{diag}[\mathbf{m}(t)]$  is the diagonal formed by the sentiment vector and note that,

$$\mathbf{m}(t) \odot d\mathbf{N}(t) = \mathbf{M}^*(t)d\mathbf{N}(t) = \text{diag}[d\mathbf{N}(t)]\mathbf{m}(t), \quad (35)$$

and, using Eq. 21,

$$d\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t) | \mathcal{H}_{t_0}] = -\omega \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t) | \mathcal{H}_{t_0}]dt + \mathbf{A}\mathbf{\Lambda} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t) | \mathcal{H}_{t_0}]dt + \omega \boldsymbol{\alpha}dt. \quad (36)$$

Then, if we substitute Eqs. 34 and 36 in Eq. 32, we obtain

$$d(\Delta \mathbf{x}) = -\omega[\mathbf{x}^*(t) - \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t) | \mathcal{H}_{t_0})]dt + \mathbf{A}[\mathbf{M}^*(t)d\mathbf{N}(t) - \mathbf{\Lambda} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t) | \mathcal{H}_{t_0})dt]. \quad (37)$$

As a result, we can write

$$\begin{aligned} d\mathbf{\Gamma}(t_0, t) &= \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}} \left[ -2\omega \Delta \mathbf{x} \Delta \mathbf{x}^T dt + \underbrace{\Delta \mathbf{x} ((\mathbf{M}^*(t)d\mathbf{N}(t) - \mathbf{\Lambda} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t) | \mathcal{H}_{t_0})dt)^T \mathbf{A}^T}_{\text{Term 1}} \right. \\ &\quad \left. + \mathbf{A}(\mathbf{M}^*(t)d\mathbf{N}(t) - \mathbf{\Lambda} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t) | \mathcal{H}_{t_0})dt) \Delta \mathbf{x}^T \right. \\ &\quad \left. + \underbrace{\mathbf{A}(\mathbf{M}^*(t)d\mathbf{N}(t) - \mathbf{\Lambda} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t) | \mathcal{H}_{t_0})dt)(\mathbf{M}^*(t)d\mathbf{N}(t) - \mathbf{\Lambda} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t) | \mathcal{H}_{t_0})dt)^T \mathbf{A}^T}_{\text{Term 2}} \right] | \mathcal{H}_{t_0}, \end{aligned} \quad (38)$$

where Term 1 gives

$$\begin{aligned} &\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}} [\Delta \mathbf{x} (\mathbf{M}^*(t)d\mathbf{N}(t) - \mathbf{\Lambda} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t) | \mathcal{H}_{t_0})dt)^T \mathbf{A}^T | \mathcal{H}_{t_0}] \\ &= \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}} [\Delta \mathbf{x} (\mathbf{m}^*(t) - \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t) | \mathcal{H}_{t_0}))^T \mathbf{A}^T | \mathcal{H}_{t_0}] \mathbf{\Lambda} dt \quad (\text{Using Eq. 35 and the fact that } \mathbb{E}[\text{diag}(d\mathbf{N}(t))] = \mathbf{\Lambda}) \\ &= \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}} [\Delta \mathbf{x} \cdot \mathbb{E}[(\mathbf{m}^*(t) - \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t) | \mathcal{H}_{t_0}))^T \mathbf{A}^T | \mathbf{x}(t), \mathcal{H}_{t_0}]] \mathbf{\Lambda} dt \\ &= \mathbf{\Gamma}(t_0, t) \mathbf{\Lambda} dt, \end{aligned}$$

and Term 2 gives

$$\begin{aligned} &\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}} \mathbf{A}[(\mathbf{M}^*(t)d\mathbf{N}(t) - \mathbf{\Lambda} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t) | \mathcal{H}_{t_0})dt)(\mathbf{M}^*(t)d\mathbf{N}(t) - \mathbf{\Lambda} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t) | \mathcal{H}_{t_0})dt)^T | \mathcal{H}_{t_0}] \mathbf{A}^T \\ &= \mathbf{A} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}} [\mathbf{M}^*(t) \mathbb{E}(d\mathbf{N}(t)d\mathbf{N}^T(t)) \mathbf{M}^*(t) | \mathcal{H}_{t_0}] \mathbf{A}^T + O(dt^2) \\ &= \mathbf{A} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}} [\mathbf{M}^2(t) | \mathcal{H}_{t_0}] \mathbf{\Lambda} dt \mathbf{A}^T \quad (\text{From Eq. 33}) \\ &= \mathbf{A}(\sigma^2 \mathbf{I} + \text{diag}(\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}(t)^2 | \mathcal{H}_{t_0})) \mathbf{\Lambda} \mathbf{A}^T dt \\ &= \mathbf{A}[\sigma^2 \mathbf{I} + \mathbf{\Gamma}_{ii}(t_0, t) + \text{diag}[\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t) | \mathcal{H}_{t_0})^2]] \mathbf{\Lambda} \mathbf{A}^T dt. \end{aligned}$$

Hence, substituting the expectations above into Eq. 38, we obtain

$$\begin{aligned} \frac{d\mathbf{\Gamma}(t_0, t)}{dt} &= -2\omega \mathbf{\Gamma}(t_0, t) + \mathbf{\Gamma}(t_0, t) \mathbf{\Lambda} \mathbf{A}^T + \mathbf{A} \mathbf{\Lambda} \mathbf{\Gamma}(t_0, t) \\ &\quad + \mathbf{A} \mathbf{\Gamma}_{ii}(t) \mathbf{\Lambda} \mathbf{A}^T + \sigma^2 \mathbf{A} \mathbf{\Lambda} \mathbf{A}^T + \mathbf{A} \text{diag}(\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t) | \mathcal{H}_{t_0}]^2) \mathbf{\Lambda} \mathbf{A}^T, \end{aligned} \quad (39)$$

Eq. 39 can be readily written in vectorial form by exploiting the properties of the Kronecker product as

$$\frac{d[\text{vec}(\mathbf{\Gamma}(t_0, t))]}{dt} = \mathbf{V}(t)\text{vec}(\mathbf{\Gamma}(t_0, t)) + \text{vec}(\sigma^2 \mathbf{A}\mathbf{\Lambda}\mathbf{A}^T + \mathbf{A} \text{diag}(\mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(t)|\mathcal{H}_{t_0}])^2 \mathbf{\Lambda}\mathbf{A}^T)$$

where

$$\mathbf{V}(t) = (-\omega \mathbf{I} + \mathbf{A}\mathbf{\Lambda}) \otimes \mathbf{I} + \mathbf{I} \otimes (-\omega \mathbf{I} + \mathbf{A}\mathbf{\Lambda}) + \sum_{i=1}^n (\mathbf{A}\mathbf{\Lambda}P_i \otimes \mathbf{A}P_i),$$

$P_i = [\delta_{ii}]$  and  $\otimes$  stands for the Kronecker product. Hence, the closed form solution of the above equation can be written as,

$$\text{vec}(\mathbf{\Gamma}(t_0, t)) = \int_0^t e^{\mathbf{V}(t-\theta)} \text{vec}[\sigma^2 \mathbf{A}\mathbf{\Lambda}\mathbf{A}^T + \mathbf{A} \text{diag}(\mathbb{E}_{\mathcal{H}(\theta) \setminus \mathcal{H}_{t_0}}[\mathbf{x}^*(\theta)|\mathcal{H}_{t_0}])^2 \mathbf{\Lambda}\mathbf{A}^T] d\theta. \quad (40)$$

Moreover, the covariance matrix  $\mathbf{\Gamma}(t_0, t)$  is bounded iff

$$\text{Re}[\lambda[(-\omega \mathbf{I} + \mathbf{A}\mathbf{\Lambda}) \otimes \mathbf{I} + \mathbf{I} \otimes (-\omega \mathbf{I} + \mathbf{A}\mathbf{\Lambda}) + (\mathbf{A} \otimes \mathbf{A})\hat{\mathbf{\Lambda}}]] < 0,$$

where  $\hat{\mathbf{\Lambda}} := \sum_{i=1}^{|\mathcal{V}|} \mathbf{\Lambda}P_i \otimes P_i$ .

In this case, the steady state solution  $\mathbf{\Gamma} = \lim_{t \rightarrow \infty} \mathbf{\Gamma}(t_0, t)$  is given by

$$-2\omega \mathbf{\Gamma} + \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{A}^T + \mathbf{A}\mathbf{\Lambda}\mathbf{\Gamma} + \mathbf{A} \text{diag}(\mathbf{\Gamma})\mathbf{\Lambda}\mathbf{A}^T + \sigma^2 \mathbf{A}\mathbf{\Lambda}\mathbf{A} + \mathbf{A} \text{diag}(\mathbb{E}_{\mathcal{H}_t}[\mathbf{x}^*])^2 \mathbf{\Lambda}\mathbf{A}^T = 0. \quad (41)$$

that means  $\lim_{t \rightarrow \infty} \text{vec}[\mathbf{\Gamma}(t_0, t)]$  is same as,

$$\begin{aligned} \text{vec}(\mathbf{\Gamma}) &= [(-\omega \mathbf{I} + \mathbf{A}\mathbf{\Lambda}) \otimes \mathbf{I} + \mathbf{I} \otimes (-\omega \mathbf{I} + \mathbf{A}\mathbf{\Lambda})^T + (\mathbf{A} \otimes \mathbf{A})\hat{\mathbf{\Lambda}}]^{-1} \\ &\quad \times \text{vec}[\sigma^2 \mathbf{A}\mathbf{\Lambda}\mathbf{A}^T + \mathbf{A} \text{diag}(\mathbf{x}^*)^2 \mathbf{\Lambda}\mathbf{A}^T] \end{aligned} \quad (42)$$

where  $\mathbf{x}^* = \lim_{t \rightarrow \infty} \mathbb{E}_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}(\mathbf{x}^*(t)|\mathcal{H}_{t_0})$ . Finally, the variance of node  $u$ 's opinion at any time  $t$  is given by

$$\sigma_{\mathcal{H}_t \setminus \mathcal{H}_{t_0}}^2(x_u^*(t)|\mathcal{H}_{t_0}) = [\mathbf{\Gamma}(t_0, t)]_{u,u} = \text{vec}(P_u)^T \text{vec}(\mathbf{\Gamma}(t_0, t)),$$

where  $P_u$  is the sparse matrix with its only  $(u, u)$  entry to be equal to unity.

## H Parameter estimation algorithm

### H.1 Estimation of $\alpha$ and $\mathbf{A}$

Algorithm H.1 summarizes the estimation of the opinion dynamics, *i.e.*,  $\alpha$  and  $\mathbf{A}$ , which reduces to a least-square problem.

---

**Algorithm 1:** Estimation of  $\alpha$  and  $\mathbf{A}$ 


---

```

1: Input:  $\mathcal{H}_T, G(V, E)$ , regularizer  $\lambda$ , error bound  $\epsilon$ 
2: Output:  $\hat{\alpha}, \hat{\mathbf{A}}$ 
3: Initialize:
4: IndexForV[1 : |V|] =  $\vec{0}$ 
5: for  $i = 1$  to  $|\mathcal{H}(T)|$  do
6:    $\mathcal{T}[u_i](\text{IndexForV}[u_i]) = (t_i, u_i, m_i)$ 
7:   IndexForV[ $u_i$ ]++
8: end for
9: for  $u \in \mathcal{V}$  do
10:   $i = 0$ 
11:   $S = \mathcal{T}[u]$ 
12:  for  $v \in \mathcal{N}(u)$  do
13:     $S = \text{MergeTimes}(S, \mathcal{T}[v])$ 
14:    // Merges two sets w.r.t  $t_i$ , takes  $O(|\mathcal{H}_u(T)| + |\cup_{v \in \mathcal{N}(u)} \mathcal{H}_v(T)|)$  steps
15:     $\mathbf{x}_{\text{last}} = 0$ 
16:     $t_{\text{last}} = 0$ 
17:     $j = 0$ 
18:    for  $i = 1$  to  $|S|$  do
19:      // This loop takes  $O(|S|) = O(|\mathcal{H}_u(T)| + |\cup_{v \in \mathcal{N}(u)} \mathcal{H}_v(T)|)$  steps
20:       $(t_v, v, m_v) = S[i]$ 
21:       $t_{\text{now}} = t_v$ 
22:      if  $u = v$  then
23:         $\mathbf{x}_{\text{now}} = \mathbf{x}_{\text{last}} e^{-\omega(t_{\text{now}} - t_{\text{last}})}$ 
24:         $j++$ 
25:         $g[u](j, v) = \mathbf{x}_{\text{now}}$ 
26:         $Y[u](j) = m_u$ 
27:      else
28:         $\mathbf{x}_{\text{now}} = \mathbf{x}_{\text{last}} e^{-\omega(t_{\text{now}} - t_{\text{last}})} + m_v$ 
29:      end if
30:       $t_{\text{last}} = t_{\text{now}}$ 
31:       $\mathbf{x}_{\text{last}} = \mathbf{x}_{\text{now}}$ 
32:    end for
33:  end for
34: end for
35: // Estimation of  $(\alpha, \mathbf{A})$ 
36: for  $u \in \mathcal{V}$  do
37:   $a = \text{InferOpinionParams}(Y[u], \lambda, g[u])$ 
38:   $\hat{\alpha}_u = a[1]$ 
39:   $\hat{\mathbf{A}}[*][u] = a[1 : \text{end}]$ 
40: end for

```

---

---

**Algorithm 2:** InferOpinionParams( $Y_u, \lambda, g_u$ )

---

```
1:  $s = \text{numRows}(g_u)$ 
2:  $X = [\mathbf{I}_s \ g_u]$ 
3:  $Y = Y_u$ 
4:  $L = X'X$ 
5:  $x = (\lambda I + L)^{-1} X'Y$ 
6: return  $x$ 
```

---

## H.2 Estimation of $(\mu, B)$

The first step of the parameter estimation procedure for temporal dynamics also involves the computation of the triggering kernels, which we do in same way as for the opinion dynamics in Algorithm H.1. In order to estimate the parameters, we adopted the spectral projected gradient descent (SPG) method proposed in [4].

---

**Algorithm 3:** Spectral projected gradient descent for  $\mu_u$  and  $B_{u^*}$ 

---

```
1: Input:  $\mathcal{H}_T, G(V, E)$  and  $\mu_u^0, B_{u^*}^0$ , step-length bound  $\alpha_{max} > 0$  and initial-step length  $\alpha_{bb} \in (0, \alpha_{max}]$ , error bound  $\epsilon$  and size of memory  $h$ 
2: Output:  $\hat{\mu}_u, \hat{B}_{u^*}$ 
3: Notation:
    $x = [\mu_u, B_{u^*}]$ 
    $f(x) = \sum_{e_i \in \mathcal{H}(T)} \log \lambda_{u_i}^*(t_i) - \sum_{u \in \mathcal{V}} \int_0^T \lambda_u^*(\tau) d\tau$ 
4: Initialize:
    $k = 0$ 
    $x_0 = [\mu_u^0, B_{u^*}^0]$ 
5: while  $\|d_k\| < \epsilon$  do
6:    $\bar{\alpha}_k \leftarrow \min\{\alpha_{max}, \alpha_{bb}\}$ 
7:    $d_k \leftarrow P_c[x_k - \bar{\alpha}_k \nabla f^u(x_k)] - x_k$ 
8:    $f_b^u \leftarrow \max\{f^u(x_k), f^u(x_{k-1}), \dots, f^u(x_{k-h})\}$ 
9:    $\alpha \leftarrow 1$ 
10:  while  $q_k(x_k + \alpha d_k) > f_b^u + \nu \alpha \nabla f^u(x_k)^T d_k$  do
11:    Select  $\alpha$  by backtracking line-search;
12:  end while
13:   $x_{k+1} \leftarrow x_k + \alpha d_k$ 
14:   $s_k \leftarrow \alpha d_k$ 
15:   $y_k \leftarrow \alpha B_k d_k$ 
16:   $\alpha_{bb} \leftarrow y_k^T y_k / s_k^T y_k$ 
17:   $k = k + 1$ 
18: end while
```

---

## I Model simulation algorithm

We leverage the simulation algorithm in [11] to design an efficient algorithm for simulating opinion-dynamics. The two basic premises of the simulation algorithm are the sparsity of the network and the Markovian property of the model. Due to sparsity, any sampled event would effect only a few number of intensity functions, only those of the local neighbors of the node. Therefore, to generate the new sample and to identify the intensity functions that require changes, we need  $O(\log |\mathcal{V}|)$  operations to maintain the heap for the priority queue. On the other hand, the Markovian property allows us to update the rate and opinion in  $O(1)$  operations. The worst-case time-complexity of this algorithm can be found to be  $O(d_{\max} |\mathcal{H}(T)| |\mathcal{V}|)$ , where  $d_{\max}$  is the maximum degree.

---

**Algorithm 4:** OpinionModelSimulation( $T, \mu, B, \alpha, A$ )

---

```
1: Initialize the priority queue  $Q$ 
2: LastOpinionUpdateTime[1 : |V|] = LastIntensityUpdateTime[1 : |V|] =  $\vec{0}$ 
3: LastOpinionUpdateValue[1 : |V|] =  $\alpha$ 
4: LastIntensityUpdateValue[1 : |V|] =  $\mu$ 
5:  $\mathcal{H}(0) \leftarrow \emptyset$ 
6: for  $u \in \mathcal{V}$  do
7:    $t = \text{SampleEvent}(\mu[u], 0, u)$ 
8:    $Q.\text{insert}([t, u])$ 
9: end for
10: while  $t < T$  do
11:   %%Approve the minimum time of all posts
12:    $[t', u] = Q.\text{ExtractMin}()$ 
13:    $t_u = \text{LastOpinionUpdateTime}[u]$ 
14:    $x_{t_u} = \text{LastOpinionUpdateValue}[u]$ 
15:    $x_{t'_u} \leftarrow \alpha[u] + (x_{t_u} - \alpha[u])e^{-\omega(t-t_u)}$ 
16:   LastOpinionUpdateTime[ $u$ ] =  $t'$ 
17:   LastOpinionUpdateValue[ $u$ ] =  $x_{t'_u}$ 
18:    $m_u \sim p(\mathbf{m}|x_u(t))$ 
19:    $\mathcal{H}(t') \leftarrow \mathcal{H}(t) \cup (t, m_u, u)$ 
20:   %% Update neighbors'
21:   for  $\forall v$  such that  $u \rightsquigarrow v$  do
22:      $t_v = \text{LastIntensityUpdateTime}[v]$ 
23:      $\lambda_{t_v} = \text{LastIntensityUpdateValue}[v]$ 
24:      $\lambda_{t'_v} \leftarrow \mu[v] + (\lambda_{t_v} - \mu[v])e^{-\omega(t-t_v)} + B_{uv}$ 
25:     LastIntensityUpdateTime[ $v$ ] =  $t'$ 
26:     LastIntensityUpdateValue[ $v$ ] =  $\lambda_{t'_v}$ 
27:      $t_v = \text{LastOpinionUpdateTime}[v]$ 
28:      $x_{t_v} = \text{LastOpinionUpdateValue}[v]$ 
29:      $x_{t'_v} \leftarrow \alpha[v] + (x_{t_v} - \alpha[v])e^{-\omega(t-t_v)} + A_{uv}m_u$ 
30:     LastOpinionUpdateTime[ $v$ ] =  $t'$ 
31:     LastOpinionUpdateValue[ $v$ ] =  $x_{t'_v}$ 
32:     %%Sample by only effected nodes
33:      $t_+ = \text{SampleEvent}(\lambda_v(t'), t', v)$ 
34:      $Q.\text{UpdateKey}(v, t_+)$ 
35:   end for
36:    $t \leftarrow t'$ 
37: end while
38: return  $\mathcal{H}(T)$ 
```

---

---

**Algorithm 5:** SampleEvent( $\lambda, t, v$ )

---

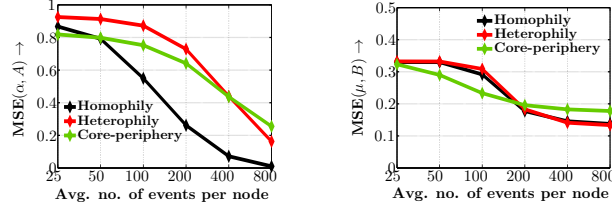
```
1:  $\bar{\lambda} = \lambda, s \leftarrow t$ 
2: while  $s < T$  do
3:    $\mathcal{U} \sim \text{Uniform}[0, 1]$ 
4:    $s = s - \frac{\ln \mathcal{U}}{\bar{\lambda}}$ 
5:    $\lambda(s) \leftarrow \mu[v] + (\lambda_v(t) - \mu[v])e^{-\omega(s-t)}$ 
6:    $d \sim \text{Uniform}[0, 1]$ 
7:   if  $d\bar{\lambda} < \lambda$  then
8:     break
9:   else
10:     $\bar{\lambda} = \lambda(s)$ 
11:   end if
12: end while
13: return  $s$ 
```

---



## J Experiments on Synthetic Data

**Parameter estimation accuracy.** We evaluate the accuracy of our model estimation procedure on three types of Kronecker networks [20]: i) Assortative networks (parameter matrix  $[0.96, 0.3; 0.3, 0.96]$ ), in which nodes tend to link to nodes with similar degree; ii) dissortative networks ( $[0.3, 0.96; 0.96, 0.3]$ ), in which nodes tend to link to nodes with different degree; and iii) core-periphery networks ( $[0.9, 0.5; 0.5, 0.3]$ ). For each network, the message intensities are multivariate Hawkes,  $\mu$  and  $B$  are drawn from a uniform distribution  $U(0, 1)$ , and  $\alpha$  and  $A$  are drawn from a Gaussian distribution  $\mathcal{N}(\mu = 0, \sigma = 1)$ . We use exponential kernels with parameters  $\omega = 100$  and  $\nu = 1$ , respectively, for opinions  $\mathbf{x}_u^*(t)$  and intensities  $\lambda^*(t)$ . We evaluate the accuracy of our estimation procedure in terms of mean squared error (MSE), between the estimated and true parameters, *i.e.*,  $\mathbb{E}[(x - \hat{x})^2]$ . Figure 5 shows the MSE of the parameters  $(\alpha, A)$ , which control the Hawkes intensities, and the parameters  $(\mu, B)$ , which control the opinion updates. As we feed more messages into the estimation procedure, it becomes more accurate.



(a) Temporal parameters,  $(\alpha, A)$  (b) Opinion parameters,  $(\mu, B)$

Figure 5: Performance of model estimation for several 512-node kronecker networks in terms of mean squared error between estimated and true parameters. As we feed more messages into the estimation procedure, the estimation becomes more accurate.

## K Twitter dataset description

We used the Twitter search API<sup>8</sup> to collect all the tweets (corresponding to a 2-3 weeks period around the event date) that contain hashtags related to the following events/topics:

- **Politics:** Delhi Assembly election, from 9th to 15th of December 2013.
- **Movie:** Release of “*Avengers: Age of Ultron*” movie, from April 28th to May 5th, 2015.
- **Fight:** Professional boxing match between the eight-division world champion Manny Pacquiao and the undefeated five-division world champion Floyd Mayweather Jr., on May 2, 2015.
- **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 5th to 16th, 2015.
- **US:** Presidential election in the United-States, from April 7th to 13th, 2016.

We then built the follower-followee network for the users that posted the collected tweets using the Twitter rest API<sup>9</sup>. Finally, 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.

Dataset	$ \mathcal{V} $	$ \mathcal{E} $	$ \mathcal{H}(T) $	$\mathbb{E}[m]$	$\text{std}[m]$
Tw: Politics	548	5271	20026	0.0169	0.1780
Tw: Movie	567	4886	14016	0.5969	0.1358
Tw: Fight	848	10118	21526	-0.0123	0.2577
Tw: Bollywood	1031	34952	46845	0.5101	0.2310
Tw: US	533	20067	18704	-0.0186	0.7135

Table 1: Real datasets statistics

## L Baselines description

- **Collab-Filter:** We adopt the proposal described by Kim *et al.* in [19] to use sentiment prediction using collaborative filtering. Let us assume that  $U$  and  $V$  are two latent low rank matrices capturing the characteristics of *spread* and *reception* of the users.  $R_{i,j}^{(t)}$  is the sentiment value posted by user  $i$  at time  $t$  that evokes a message of user  $j$  before any other neighbor of  $j$  posts a message. To compute  $U$  and  $V$ , we minimize the following optimization problem.

$$\sum_{t \in \mathcal{H}_t} \sum_{i,j \in \mathcal{V}} I_{i,j} (R_{i,j}^{(t)} - U_i^T V_j)^2 + \lambda_1 \sum_i \|U_i\|_2^2 + \lambda_2 \sum_j \|V_j\|_2^2$$

Here  $I_{i,j}$ ’s are the entries of adjacency matrix.

- **Flocking:** In the flocking model [15], a node  $i$  with opinion  $x$  first selects the set of neighbors  $j$  having opinions  $x_j$  so that  $|x_i - x_j| < \epsilon$  and then updates her own opinion by averaging them.
- **Voter:** In this strategy [26], at each step, a node is selected at random; she chooses one of its neighbors uniformly at random (including itself) and adopts that opinion as its own.
- **DeGroot:** This model [9] suggests that a node updates her opinion by taking a weighted average of her neighbors’ opinion. In particular, this proposal assumes that the array of weights form a row-stochastic influence matrix with  $w_{i,j} \geq 0$ , and opinions in the range  $[0, 1]$  (which stochastic updates preserved).
- **Linear:** A generalize linear model [8] generalizes the DeGroot model by relaxing the structure of influence matrix. Under this model, the influence matrix need not be stochastic. Moreover, the influence parameters can take negative values.
- **Biased Voter:** A biased voter model [7] captures various aspects of other opinion models by a mix of three factors: stubbornness (ignoring others opinions), DeGroots weighted averaging with neighbors, and biased conformance (Flocking Model), which selects users having opinion close to that of the base agent.

<sup>8</sup><https://dev.twitter.com/rest/public/search>

<sup>9</sup><https://dev.twitter.com/rest/public>