

# Modeling Temporal Activity Patterns in Dynamic Social Networks

Vasanthan Raghavan<sup>\*1</sup>, Greg Ver Steeg<sup>2</sup>, Aram Galstyan<sup>2</sup>, and Alexander G. Tartakovsky<sup>1</sup>

<sup>1</sup>Department of Mathematics, University of Southern California, Los Angeles, 90089, CA, USA

<sup>2</sup>Information Sciences Institute, University of Southern California, Marina del Rey, 90292, CA, USA

Email: Vasanthan Raghavan<sup>\*</sup> - vasanthan\_raghavan@ieee.org; Greg Ver Steeg - gregv@isi.edu; Aram Galstyan - galstyan@isi.edu; Alexander G. Tartakovsky - tartakov@usc.edu;

<sup>\*</sup>Corresponding author

## Abstract

The focus of this work is on developing probabilistic models for user activity in social networks by incorporating the social network *influence* as perceived by the user. For this, we propose a *coupled* Hidden Markov Model, where each user's activity evolves according to a Markov chain with a hidden state that is influenced by the collective activity of the friends of the user. We develop generalized Baum-Welch and Viterbi algorithms for model parameter learning and state estimation for the proposed framework. We then validate the proposed model using a significant corpus of user activity on Twitter. Our numerical studies show that with sufficient observations to ensure accurate model learning, the proposed framework explains the observed data better than either a renewal process-based model or a conventional uncoupled Hidden Markov Model. We also demonstrate the utility of the proposed approach in predicting the time to the next tweet. Finally, clustering in the model parameter space is shown to result in distinct natural clusters of users characterized by the interaction dynamic between a user and his network.

## Keywords

Activity Profile Modeling, Twitter, Data-Fitting, Explanation, Prediction, Hidden Markov Model, Coupled Hidden Markov Model, Social Network Influence, User Clustering

## Introduction

Social networking websites such as **Facebook**, **Twitter**, etc. have become immensely popular with hundreds of millions of users that engage in various forms of activity on these websites. These social networks provide an unparalleled opportunity to study individual and collective behavior at a very large scale. Such studies have profound implications on wide-ranging applications such as efficient resource allocation, user-specific information dissemination, user classification, and rapid detection of anomalous behavior such as bot or compromised accounts, etc.

The simplest model for user activity is a Poisson process, where each activity event (e.g., posting, tweeting, etc.) occurs independent of the past history at a time-independent rate. However, recent empirical evidence from multiple sources (e-mail logs, web surfing, letter correspondence, research output, etc.) suggest that human activity has distinctly non-Poissonian characteristics. In particular, the inter-event duration distribution (which is exponential for the Poisson process) has been shown to be heavy-tailed and bursty for a number of different activity types [1, 2]. Different approaches have been put forward to explain the non-Poisson nature of the activity patterns [3–10]. Nevertheless, despite significant recent progress, open questions remain. Most remarkably, on an individual scale, existing studies so far have mostly discarded the role and impact of the social network where the user activity takes place, instead describing each user via an *independent* stochastic process. On the other hand, it is clear that social interactions on networks affect user activity, and discarding these interactions should generally lead to sub-optimal models.

The main contribution of this paper is to develop a probabilistic model of user activity that explicitly takes into account the interaction between users by introducing a coupling between two stochastic processes. Specifically, we propose a coupled Hidden Markov Model (coupled HMM) to describe inter-connected dynamics of user activity. In our model, the individual dynamics of each user is coupled to the aggregated activity profile of his neighbors (friends or followers) in the network. While a user’s activity may be preferentially affected by specific neighbors, the predictive power of the model can be substantially improved using the aggregated activity of all the neighbors. The hidden states in our model correspond to different patterns in user activity, similar to the approach suggested in [7]. However, here the state transitions are influenced by the activity of the neighbors, and in turn, the activity of the aggregated set of neighbors is influenced by the state of the given user. While many variants of the classical HMM approach exist in the literature, the key distinguishing feature of this work is the *bi-directional* influence between the activity of a user and his network, specifically tailored for social network applications. Nevertheless, being a variant of the conventional (uncoupled) HMM, the proposed model enjoys the same computational advantages in

terms of parameter learning as [7] and other HMM variants.

We perform a number of experiments with data describing user activity traces on **Twitter**, and demonstrate that the proposed approach has a better performance both in terms of explaining observed data (model fitting) and predicting future activity (generalization). In particular, we report statistically significant improvement over two baseline approaches, a renewal process-based model and a conventional HMM. Furthermore, we use the learned models to cluster users, and find that the resulting cluster structure allows intuitive characterization of the users in terms of the interaction dynamics between a user and his social network.

The rest of the paper is organized as follows. Section 2 reviews related work in activity profile modeling, especially as applicable in social network settings. Section 3 proposes a coupled HMM framework for the activity profile of users and distinguishes the proposed model from prior work. Section 4 develops statistical methodologies to learn model parameters and how to use the proposed framework for data-fitting and forecasting. Section 5 validates the modeling assumptions and illustrates the utility and efficacy of the proposed approach with data from [11]. Concluding remarks are provided in Section 6. Without any prejudice, we use the male gender-specific connotations for all the users and their attributes in this work.

## **Related Work**

Recent research on social networks has focused on understanding the properties of networks induced by social interactions, modeling information diffusion on such networks, characterizing their evolution in time, etc. In the direction of modeling the temporal activity of users' communication in such networks, several models have been proposed in the literature. Approaches based on simple point process models have been proposed for user requests in a peer-to-peer setting [12] and social network evolution [13]. In addition to one-parameter exponential observation density for user activity utilized in [6], more general two-parameter models such as the Weibull (or stretched exponential) have been proposed for modeling inter-post duration in the context of instant-messaging networks [14], accessing patterns in Internet-media [15], understanding inter-post dynamics of original content in general online social networks [16], and inter-call duration in cell-phone networks [17].

To explain the bursty features of human dynamics, [1–4] suggested the priority selection/queue mechanism. An alternative mechanism motivated by circadian and weekly cycles of human activity and captured by cascading non-homogenous Poisson processes was suggested in [5, 6] for the heavy-tails in inter-event durations. Although this model has been shown to be consistent with empirical observations, it is computa-

tionally intensive in terms of parameter estimation. To overcome this issue, Malmgren *et al.* [7] suggested a simpler two-state HMM for the activity of users in an email/communication network where the states reflect a measure of the user’s activity. Similar models, suggestive of a few states determining activity patterns, have been considered for short message correspondence in [18] and Digg activity in [19]. Other work has also emphasized the importance of distinguishing active versus inactive users in activity and influence modeling tasks [20].

An important feature that characterizes the priority selection mechanism is that human behavioral patterns are driven by responses to/from others. On the other hand, the line of work motivated by the cascading non-homogenous Poisson processes-based models explains activity patterns due to other mechanisms such as circadian and weekly cycles, task repetition, and changing communication needs. Models that traverse between these two extremes and incorporate the influence of a user’s social network on activity have not been studied in extensive detail in the literature. Some of the examples of such a bridging effort include user-generated activity traces used for inferring underlying social relationships [21–23], incorporating similarities in the behavior of users in a specific user’s social network to model his actions [24, 25], a Poisson regression model to determine the users that most influence a given user [26], and a user activity-driven (rather than connectivity-driven) model for social network evolution [27]. Notwithstanding the fact that similar ideas have been sporadically pursued in other settings, our work provides the missing link for a network-driven approach to capturing individual behavioral patterns.

While the theory of classical HMMs is well-developed [28, 29], HMMs are ill-suited in settings where multiple processes interact with each other and/or information about the history of the process needed for future inferencing is not reflected in the current state. To incorporate complicated dependencies between several interacting variables, many variants of the classical HMM set-up that are special cases of the more general theory of *dynamic Bayesian networks* [30, 31] have been introduced in the literature. Some of these extensions include the class of *autoregressive* HMMs [32–34], *input-output* HMMs [35, 36], *factorial* HMMs [37], and *coupled* HMMs. Different variants of coupled HMMs have been used in diverse settings including models for complex human actions and behaviors [38, 39], freeway traffic [40], audio-visual speech [41, 42], EEG classification [43], spread of infection in social networks [44], etc.

## Modeling Activity Profile of Twitter Users

Let  $T_i, i = 0, 1, \dots, N$  denote the time-stamps of a specific user's tweets over a given period-of-interest. We can equivalently define the inter-tweet duration  $\Delta_i$  as

$$\Delta_i \triangleq T_i - T_{i-1}, \quad i = 1, 2, \dots, N.$$

One of the main goals of this work is to develop a mathematical model for  $\{\Delta_i\} \triangleq \Delta_1^N = [\Delta_1, \dots, \Delta_N]$ . Along the lines of [7], we start by developing a simplistic  $k = 2$ -state HMM for  $\{\Delta_i\}$ .

### Influence-Free Hidden Markov Modeling

**Assumption 1 – Underlying States:** We assume that a variable  $Q_i$ , taking one of two possible values  $\{0, 1\}$ , reflects the state of the user-of-interest. Specifically,  $Q_i = 0$  denotes that the user is in an *Inactive* state between  $T_{i-1}$  and  $T_i$ , whereas  $Q_i = 1$  denotes that the user is in an *Active* state. We also assume that  $Q_i$  ( $i \geq 1$ ) evolves in a time-homogenous Markovian manner and is dependent only on  $Q_{i-1}$  and is conditionally independent of  $Q_0^{i-2} = [Q_0, \dots, Q_{i-2}]$  given  $Q_{i-1}$ . A first-order Markovian model is a reasonable first approximation capturing short-term memory in human behavioral dynamics [2, 45]. The state transition probability matrix  $\mathbf{P} = \{P[m, n]\}$  is given as

$$\mathbf{P} = \begin{bmatrix} 1 - \beta_{0,1} & \beta_{0,1} \\ \beta_{1,0} & 1 - \beta_{1,0} \end{bmatrix},$$

with  $P[m, n] = P(Q_i = n | Q_{i-1} = m)$ ,  $m, n \in \{0, 1\}$ . The density of the initial state  $Q_0$  is denoted as  $P(Q_0 = j) = \pi_j$ ,  $j = 0, 1$ . Note that the switching from the *Inactive* state to the *Active* state in the HMM paradigm can capture the nocturnal/work-home patterns of individual users without any further explicit modeling [7]. Further, explicit modeling of circadian and weekly cycles in social network settings is more difficult than for email communications due to the “often on” more random nature of social network interactions.

**Assumption 2 – Observation Density:** In general,  $Q_i$  is hidden (unobservable) and we can only observe  $\{\Delta_i\}$  (or equivalently,  $\{T_i\}$ ). In the *Inactive* state,  $\{\Delta_i\}$  form samples from a “low”-rate point process, whereas in the *Active* state,  $\{\Delta_i\}$  form samples from a “high”-rate point process. Specifically, let the probability density function of  $\Delta_i$  be given as

$$\Delta_i \sim \begin{cases} f_1(\cdot) & \text{if } Q_i = 1 \\ f_0(\cdot) & \text{if } Q_i = 0, \end{cases}$$

for an appropriate choice of  $f_0(\cdot)$  and  $f_1(\cdot)$ .

As mentioned earlier, an exponential model for  $f(\cdot)$  corresponds to a Poisson process assumption under either state. While the exponential model is captured by a single parameter (see Table 1 for details), this simplicity often constrains the model fit either in the small inter-tweet (bursty) regime or large inter-tweet regime (tails). Two-parameter extensions of the exponential such as the gamma or Weibull density allow a better fit in these two regimes. Both the gamma and the Weibull models result in similar modeling fits. While the gamma model allows for simple parameter estimate formulas (see Table 1), the Weibull model requires the solving of coupled equations in the model parameters (often, a numerically intensive procedure). Thus, we will restrict attention to the exponential and gamma model choices in this work.

### Influence-Driven Hidden Markov Modeling

A more sophisticated influence-driven model is developed now by making the following additional assumptions:

**Assumption 3 – Influence of Neighbors:** In addition to  $Q_{i-1}$ , the evolution of  $Q_i$  is also influenced by the aggregated activity of all the users interacting with and influencing the user-of-interest (“neighbors,” for short). While neighbors is a broad rubric, we restrict attention to the friends and followers in this work. For example, a series of tweets from the neighbors can result in a reply/retweet by the user, or a long period of non-activity from the neighbors could induce the user to initiate a burst of activity. Let the variable  $Z_i$  ( $i = 1, \dots, N$ ) capture the influence of the neighbors’ tweets on the user-of-interest. Examples of candidate influence structures (information theoretically,  $Z_i$  is viewed as a side-information metric) include:

1. A binary indicator function that reflects whether there was a mention of the user between  $T_{i-1}$  and  $T_i$  (or not);
2. The number of such mentions;
3. Total traffic (aggregated activity) of the friends of the user, etc.

Motivated by the same short-term memory assumption as before, the *coupling* between  $\{Q_i\}$  and  $\{Z_i\}$  is simplified by the Markovian condition that  $P(Q_i|Q_1^{i-1}, Z_1^i) = P(Q_i|Q_{i-1}, Z_i)$ . In general, to keep computational requirements in inferencing low, it is helpful to assume that the evolution of  $Q_i$  is captured by a summary statistic  $\phi(Z_i) : Z_i \mapsto [0, 1]$  such that

$$P(Q_i|Q_{i-1}, Z_i) = P_0(Q_i|Q_{i-1}) \cdot (1 - \phi(Z_i)) + P_1(Q_i|Q_{i-1}) \cdot \phi(Z_i)$$

with  $P_k[m, n] = P_k(Q_i = n | Q_{i-1} = m)$  where

$$P_0 = \begin{bmatrix} 1-p_0 & p_0 \\ q_0 & 1-q_0 \end{bmatrix} \quad \text{and} \quad P_1 = \begin{bmatrix} 1-p_1 & p_1 \\ q_1 & 1-q_1 \end{bmatrix}.$$

In particular, the choice  $\phi(Z_i) = \mathbb{1}(Z_i > \tau)$  for a suitable threshold  $\tau$  implies that the user switches from the transition probability matrix  $P_0$  to  $P_1$  depending on the magnitude of the influence structure. To paraphrase, the user evolves according to a baseline dynamics corresponding to  $P_0$  if his network activity is below a certain threshold and evolves according to an elevated dynamics corresponding to  $P_1$  if his network activity exceeds that threshold. The discussion on model learning elaborates on a simple method to determine the appropriate choice of  $\tau$ . The discussion on model validation provides some empirical justification for this assumption.

**Assumption 4 – Evolution of Influence Structure:** Noting that  $Z_i$  is a function of the activity of all the neighbors (and not a specific user), we hypothesize that  $Z_i$  is dependent on  $Q_{i-1}$ , but only weakly dependent on  $Z_{i-1}$ . Motivated by this thinking, we make the simplistic assumption that

$$P(Z_i | Q_1^{i-1}, Z_1^{i-1}) = P(Z_i | Q_{i-1}). \quad (1)$$

Rephrasing, (1) presumes that user aggregation de-correlates  $Z_i$  from its past history. While the above assumption can be justified under certain scenarios (see the discussion under model validation), more general influence evolution models need to be considered and the loss in explanatory/predictive power by making the simplistic assumption in (1) needs to be studied carefully. This is the subject of ongoing work. Further, let the probability density function of  $Z_i$  be given as

$$Z_i \sim \begin{cases} g_0(\cdot) & \text{if } Q_{i-1} = 0 \\ g_1(\cdot) & \text{if } Q_{i-1} = 1. \end{cases}$$

Different candidates for  $g_j(\cdot)$  are illustrated in Table 2.

Combining the above four assumptions, the joint density of the observations  $\{\Delta_i\}$ , the influence structure  $\{Z_i\}$ , and the state  $\{Q_i\}$  can be simplified as

$$P(\Delta_1^N, Z_1^N, Q_0^N) = P(Q_0, Z_1, Q_1, \Delta_1, \dots, Z_N, Q_N, \Delta_N) \quad (2)$$

$$= P(Q_0) \prod_{i=1}^N P(Z_i | Q_{i-1}) \prod_{i=1}^N P(Q_i | Q_{i-1}, Z_i) \prod_{i=1}^N P(\Delta_i | Q_i). \quad (3)$$

The dependence relations that drive the coupled HMM framework for user activity are illustrated in Figs. 1 and 2.

### Comparison with Related Models and Architectures

Many extensions to the classical HMM architecture have been proposed in the literature for inferencing problems in different settings. We now compare the proposed coupled HMM architecture with some of these extensions and variants. The conditional dependencies of the involved variables corresponding to the different architectures relative to the proposed model in this work are summarized in Table 3. The number of model parameters for these architectures with  $k$  states,  $\ell$  network influence structure levels, and  $m$  parameters for the observations are presented in Table 4.

The simplest extension of the HMM architecture, an *autoregressive* HMM [33], ties the evolution of the user’s inter-tweet duration to his state and his past observations. Thus, this model does not incorporate the influence of the user’s network on his activity. A more general framework called a class of *input-output* HMMs is proposed in [35] and [36], where the user’s inter-tweet duration is not only dependent on his state, but also on an *external* input such as the network influence structure. In contrast, a coupled variation of the *factorial* HMM in [37] takes the viewpoint of the network influence structure being an additional state rather than an external input. In either model, the current state of the user depends not only on his past state, but also on the state of his network. However, in both models, the evolution of the user’s network is *one-sided* and independent of the user’s interaction.

The case of a *fully-coupled* HMM overcomes this one-sided evolution by ensuring that the state of the user and his network *influence* each other. The price to pay for such generality (lack of structure in the conditional dependencies) is that the model parameters have to be learned via approximation algorithms instead of iterative techniques. To overcome this difficulty, a structured architecture is proposed in [38], where

$$\begin{aligned} P(Q_i | \{Q_{i-1}, Z_{i-1}\}) &= P(Q_i | Q_{i-1}) \cdot P(Q_i | Z_{i-1}) \\ P(Z_i | \{Q_{i-1}, Z_{i-1}\}) &= P(Z_i | Q_{i-1}) \cdot P(Z_i | Z_{i-1}). \end{aligned}$$

Similarly, [43] proposes another architecture, where

$$\begin{aligned} P(Q_i | \{Q_{i-1}, Z_{i-1}\}) &= \beta_1 \cdot P(Q_i | Q_{i-1}) + \beta_2 \cdot P(Q_i | Z_{i-1}) \\ P(Z_i | \{Q_{i-1}, Z_{i-1}\}) &= \gamma_1 \cdot P(Z_i | Q_{i-1}) + \gamma_2 \cdot P(Z_i | Z_{i-1}) \end{aligned}$$

for appropriate normalization constants  $\{\beta_i\}$  and  $\{\gamma_i\}$ . While parameter learning algorithms simplify in either scenario, from a social network perspective, the number of model parameters remain large for small values of  $k$  and  $\ell$  relative to the proposed coupled HMM architecture in this work. For example, these

structured coupled HMMs are described by 10 and 18 model parameters with an  $m = 1$  parameter observation density in each state (and 12 and 20 parameters with  $m = 2$ ) relative to 8 and 10 model parameters with the proposed model in the same settings (see Table 4 for details).

In the backdrop of this discussion, the proposed coupled HMM architecture offers a principled, novel and easily-motivated modeling framework, specifically useful in social network contexts. Despite being simple, it offers significant performance gains over the classical HMM architecture and its variants. As the subsequent discussion also shows, the proposed architecture has the added benefit of allowing model learning via simple re-estimation formulas.

## Methodology

### Learning Model Parameters

It is of interest to infer the underlying states  $\{Q_i\}$  that cannot be observed directly. This task is performed with the aid of the observations  $\{\Delta_i\}$  in the HMM setting, and with the aid of  $\{\Delta_i\}$  and the influence structure  $\{Z_i\}$  in the coupled HMM setting.

In the HMM setting, a *locally* optimal choice of model parameters is sought to maximize the likelihood function  $P(\Delta_1^N | \lambda)$ . Following the result in [46, 47], starting with an initial choice of HMM parameters  $\bar{\lambda}$ , the model parameters are re-estimated to maximize Baum’s auxiliary function  $Q(\lambda, \bar{\lambda})|_{\text{HMM}}$ , defined as,

$$Q(\lambda, \bar{\lambda})|_{\text{HMM}} \triangleq \sum_{Q_0^N} \log \left( P(\Delta_1^N, Q_0^N | \lambda) \right) \cdot P(\Delta_1^N, Q_0^N | \bar{\lambda}).$$

It can be easily checked [28, 29] that this maximization breaks into a term-by-term optimization of individual model parameters. The model re-estimation formulas are given by the Baum-Welch algorithm:

$$\hat{\beta}_{i,j} = \frac{\sum_{i=1}^N \xi_{i-1}(i,j)}{\sum_{i=1}^N (\xi_{i-1}(i,0) + \xi_{i-1}(i,1))}, \quad i \neq j, \quad i, j \in \{0, 1\}.$$

The Baum-Welch estimate of the parameters defining the different observation densities are presented in Table 1. In these equations,  $\xi_i(a, b)$  for  $i = 0, \dots, N - 1$  is defined as

$$\xi_i(a, b) \triangleq P(Q_i = a, Q_{i+1} = b | \Delta_1^N, \bar{\lambda}).$$

The update equation for  $\xi_i(a, b)$  follows from [28, Eq. (37)] and the forward-backward procedure.

In the coupled HMM setting, as in [36], we are interested in model parameters that maximize the *conditional* likelihood function  $P(\Delta_1^N | Z_1^N, \lambda)$ . If the conditional likelihood is known in closed-form (as in the input-output HMM case [36]) or a tight lower bound to it is known, a conditional expectation maximization

algorithm along the lines of [48–50] can be pursued. In the proposed coupled HMM setting, the conditional likelihood appears to be neither amenable to a simple formula nor a tight lower bound. To overcome this technical difficulty, we now propose a two-step procedure to learn an estimate of the model parameters that maximize  $P(\Delta_1^N | Z_1^N, \boldsymbol{\lambda})$ .

In the first step, we fix the threshold that distinguishes the baseline dynamics  $P_0$  from the elevated dynamics  $P_1$ ,  $\tau$ , to an appropriate choice  $\boldsymbol{\tau}_{\text{init}}$ . We then treat  $\Delta_1^N$  and  $Z_1^N$  as training observations and consider a generalized auxiliary function  $Q(\boldsymbol{\lambda}, \bar{\boldsymbol{\lambda}}) \Big|_{\text{CHMM}}$  of the form:

$$Q(\boldsymbol{\lambda}, \bar{\boldsymbol{\lambda}}) \Big|_{\text{CHMM}} \triangleq \sum_{Q_0^N} \log \left( P(\Delta_1^N, Z_1^N, Q_0^N | \boldsymbol{\lambda}) \right) \cdot P(\Delta_1^N, Z_1^N, Q_0^N | \bar{\boldsymbol{\lambda}}). \quad (4)$$

As before,  $\boldsymbol{\lambda}$  and  $\bar{\boldsymbol{\lambda}}$  denote the optimization variable corresponding to the parameter space (all the parameters except for  $\tau$ ) and its initial estimate, respectively. A straightforward extension of the proof in [46, 47] shows that maximizing  $Q(\boldsymbol{\lambda}, \bar{\boldsymbol{\lambda}}) \Big|_{\text{CHMM}}$  in the  $\boldsymbol{\lambda}$  variable results in a local maximization of the *joint* likelihood function  $P(\Delta_1^N, Z_1^N | \boldsymbol{\lambda})$ . To obtain analogous re-estimation formulas for an iterative solution to a local maximum, we define the equivalent intermediate variable  $\tilde{\xi}_i(a, b)$  for  $i = 0, \dots, N - 1$ :

$$\tilde{\xi}_i(a, b) \triangleq P(Q_i = a, Q_{i+1} = b | \Delta_1^N, Z_1^N, \bar{\boldsymbol{\lambda}}).$$

Using (3), we can simplify (4) and the joint optimization of the model parameters again breaks into a term-by-term optimization. We then have the following analogous model parameter estimates for  $k \in \{0, 1\}$ :

$$\begin{aligned} \tilde{p}_k &= \frac{\sum_{i=1, i \in \mathcal{Z}_k}^N \tilde{\xi}_{i-1}(0, 1)}{\sum_{i=1, i \in \mathcal{Z}_k}^N \tilde{\xi}_{i-1}(0, 0) + \sum_{i=1, i \in \mathcal{Z}_k}^N \tilde{\xi}_{i-1}(0, 1)} \\ \tilde{q}_k &= \frac{\sum_{i=1, i \in \mathcal{Z}_k}^N \tilde{\xi}_{i-1}(1, 0)}{\sum_{i=1, i \in \mathcal{Z}_k}^N \tilde{\xi}_{i-1}(1, 0) + \sum_{i=1, i \in \mathcal{Z}_k}^N \tilde{\xi}_{i-1}(1, 1)} \end{aligned}$$

with  $\mathcal{Z}_0 = \{i : Z_i \leq \tau\}$  and  $\mathcal{Z}_1 = \{i : Z_i > \tau\}$ . The re-estimation formulas for the observation density parameters follow the same structure as in Table 1 by replacing  $\xi_i(\cdot, \cdot)$  with  $\tilde{\xi}_i(\cdot, \cdot)$ . For the parameters defining the density of the influence structure, Table 2 provides a list of re-estimation formulas. The intermediate variable  $\tilde{\xi}_i(a, b)$ ,  $1 \leq i \leq N - 1$  is updated by a generalized forward-backward procedure whose steps are illustrated in Table 5 (see (7)-(13)).

We denote by  $\boldsymbol{\lambda}_{\text{init}}$  the converged model parameters that locally maximize  $Q(\boldsymbol{\lambda}, \bar{\boldsymbol{\lambda}}) \Big|_{\text{CHMM}}$ . Note that  $\boldsymbol{\lambda}_{\text{init}}$  is a local maximum only in the  $\boldsymbol{\lambda}$  space and not in  $\{\boldsymbol{\tau} \times \boldsymbol{\lambda}\}$ . Thus, the choice  $\{\boldsymbol{\tau}_{\text{init}}, \boldsymbol{\lambda}_{\text{init}}\}$  does not maximize  $P(\Delta_1^N | Z_1^N, \boldsymbol{\lambda})$ , not even locally. Therefore, in the next step, we locally optimize the conditional likelihood

over a local region around  $\{\boldsymbol{\tau}_{\text{init}}, \boldsymbol{\lambda}_{\text{init}}\}$ . That is,

$$\{\tilde{\boldsymbol{\tau}}, \tilde{\boldsymbol{\lambda}}\} = \arg \max_{\{\boldsymbol{\tau}, \boldsymbol{\lambda}\} \in \mathcal{L}} \mathbb{P}(\Delta_1^N | Z_1^N, \boldsymbol{\lambda})$$

where  $\mathcal{L} = \{\boldsymbol{\tau} : \boldsymbol{\tau} = \boldsymbol{\tau}_{\text{init}} + \Delta\boldsymbol{\tau} \text{ and } \boldsymbol{\lambda} : \boldsymbol{\lambda} = \boldsymbol{\lambda}_{\text{init}} + \Delta\boldsymbol{\lambda}\}$ . In this work, we focus on a box-constrained  $\mathcal{L}$ . Alternately, a local gradient search in the model parameter space can be pursued to locally optimize the conditional likelihood function. Note that the conditional density can be written as

$$\mathbb{P}(\Delta_1^N | Z_1^N, \boldsymbol{\lambda}) = \frac{\tilde{\alpha}_N(0) + \tilde{\alpha}_N(1)}{\hat{\alpha}_N(0) + \hat{\alpha}_N(1)},$$

where the forward algorithm variable  $\tilde{\alpha}_i(j)$  is updated with the same formulas as (8)-(10) (see Table 5). On the other hand,  $\hat{\alpha}_i(j)$  follows the same formula as  $\tilde{\alpha}_i(j)$ , but by constraining  $\mathbb{P}(\Delta_i | Q_i = a) = 1$  for all  $i$  and  $a$ .

### Model Verification

The efficacy of the different models to the observed data are studied in two ways. In the first approach, the model parameters learned via the (generalized) Baum-Welch algorithm are used with a state estimation procedure to estimate the most probable state sequence associated with the observations. For the HMM setting, state estimation is straightforward via the use of the Viterbi algorithm [28]. The generalization of the Viterbi algorithm to the coupled HMM setting requires the definition of intermediate variables  $\delta_{i+1}(j)$  and  $\phi_{i+1}(j)$  as illustrated in Table 5 (see (14)-(19)). As with the Viterbi algorithm, the most probable state sequence is then estimated as

$$\begin{aligned} Q_N^* &= \arg \max_{j \in \{0,1\}} \delta_N(j) \\ Q_i^* &= \phi_{i+1}(Q_{i+1}^*), \quad 1 \leq i \leq N-1. \end{aligned}$$

The observed inter-tweet durations corresponding to the classified states are compared with the inter-tweet durations obtained with the proposed model(s) via a graphical method such as the Quantile-Quantile (Q-Q) plot. Recall that a Q-Q plot plots the quantiles corresponding to the true observations with the quantiles corresponding to the model(s) [51]. If the proposed model reflects the observations correctly, the quantiles lie on the (reference) straight-line that extrapolates the first and the third quartiles. Discrepancies from the straight-line benchmark indicate artifacts introduced by the model(s) not seen in the observations and/or features in the observation not explained by the model(s).

In the second approach, the fits of the different models to the data are studied via a more formal metric such as the Akaike Information Criterion (AIC), defined as

$$\text{AIC}(n) \triangleq 2k - 2\log(\mathbf{L}),$$

where  $k$  denotes the number of parameters used in the model,  $n$  is the length of the observation sequence, and  $\mathbf{L}$  is the optimized likelihood function for the observation sequence corresponding to the learned model. The AIC penalizes models with more parameters and the model that results in the smallest value of AIC is the most suitable model (for the observed data) from the class of models considered. In the HMM setting with  $k_{\text{HMM}}$  parameters, the AIC corresponding to  $\{\Delta_i\}$  is given as

$$\begin{aligned} \text{AIC}(n) \Big|_{\text{HMM}} &\triangleq 2k_{\text{HMM}} - 2\log(\mathbf{P}(\Delta_1^N | \boldsymbol{\lambda})) \\ &= 2k_{\text{HMM}} - 2\log(\alpha_N(0) + \alpha_N(1)), \end{aligned}$$

where the converged model parameter estimates from the Baum-Welch algorithm are used to compute  $\alpha_i(j) \triangleq \mathbf{P}(\Delta_1^i, Q_i = j)$  using the forward procedure. In the coupled HMM setting with  $k_{\text{CHMM}}$  parameters, the corresponding AIC metric is defined as

$$\text{AIC}(n) \Big|_{\text{CHMM}} \triangleq 2k_{\text{CHMM}} - 2\log(\mathbf{P}(\Delta_1^N | Z_1^N, \boldsymbol{\lambda})), \quad (5)$$

where the model parameter estimates that maximize  $\mathbf{P}(\Delta_1^N | Z_1^N, \boldsymbol{\lambda})$  are to be used in (5). With the model parameters learned as explained in the previous section, an upper bound to  $\text{AIC}(n) \Big|_{\text{CHMM}}$  is obtained as

$$\text{AIC}(n) \Big|_{\text{CHMM}} \leq \overline{\text{AIC}}(n) \triangleq 2k_{\text{CHMM}} - 2\log\left(\frac{\tilde{\alpha}_N(0) + \tilde{\alpha}_N(1)}{\hat{\alpha}_N(0) + \hat{\alpha}_N(1)}\right).$$

## Forecasting

Given  $\Delta_1^n$  (and  $Z_1^n$ ), forecasting  $\Delta_{n+1}$  is of immense importance in tasks such as advertising, anomaly detection (detecting when a compromised account will post next), etc. A simple maximum *a posteriori* (MAP) predictor of the form

$$\tilde{\Delta}_{n+1} \Big|_{\text{MAP}} = \arg \max_y f(\Delta_{n+1} = y | \Delta_1^n, Z_1^n) = \arg \max_y \sum_{i=0}^{k-1} \tilde{\beta}_i f_i(\Delta_{n+1} = y)$$

where

$$\tilde{\beta}_i = \frac{\sum_j \tilde{\alpha}_n(j) \mathbf{P}(Z_{n+1} | Q_n = j) \mathbf{P}(Q_{n+1} = i | Z_{n+1}, Q_n = j)}{\sum_j \tilde{\alpha}_n(j)}$$

fails when  $f_i(\Delta_{n+1} = y)$  is unimodal with the same mode for all  $i$ . This is always the case with exponential observation models (mode is 0) and with gamma models if  $k_i \lambda_i < 1$  for all  $i$  (mode is 0), which is typically the case with the best model fits for many users. On the other hand, a conditional mean predictor of the form

$$\tilde{\Delta}_{n+1} \Big|_{\text{CM}} = \mathbb{E}[\Delta_{n+1} | \Delta_1^n, Z_1^n] = \sum_{i=0}^{k-1} \tilde{\beta}_i \mathbb{E}[\Delta_{n+1} | Q_{n+1} = i]$$

results in large forecasting errors in the *Inactive* state if the mean inter-tweet durations in the two states are very disparate (typically the case for many users). To overcome these problems, we consider a predictor of the form

$$\tilde{\Delta}_{n+1} = \sum_{i=0}^{k-1} \mathbb{1}(\tilde{Q}_{n+1} = i | \Delta_1^n) \mathbb{E}[\Delta_{n+1} | Q_{n+1} = i]$$

where  $\tilde{Q}_{n+1}$  is the state estimate using the (generalized) Viterbi algorithm with  $\Delta_1^n$  (and  $Z_1^n$ ) as inputs and study the forecasting performance in the *Active* state with a Symmetric Mean Absolute Percentage Error (SMAPE) metric:

$$\text{SMAPE}(N) \triangleq \frac{1}{N} \sum_{i=1}^N \left| \frac{\Delta_i - \tilde{\Delta}_i}{\Delta_i + \tilde{\Delta}_i} \right| \cdot \mathbb{1}(Q_i = 1).$$

The SMAPE metric is a normalized error metric and a smaller value indicates a better model for forecasting. It is seen as a percentage error and is bounded between 0% and 100%.

## Numerical Results

The dataset used to illustrate the efficacy of the models proposed in this work is a 30-day long record of **Twitter** activity described in [11]. This dataset consists of  $\mathcal{N}_t = 652,522$  tweets from  $\mathcal{N}_u = 30,750$  users (with at least one tweet). The time-scale on which the tweets are collected is minutes. While the dataset has been collected using a snowball sampling technique and reflects a population primarily based out of West Asia, London and Pakistan, the users in the dataset appear to be from diverse socio-economic and political backgrounds and have a broad array of interests. Further, the properties of the dataset on a collective scale are similar in nature to well-understood properties of similar datasets [11] suggesting a high confidence on the suitability of the dataset in studying user behavior on an individual scale and in its generalizability to other datasets.

Since reliable model learning can be accomplished only for users with sufficient activity, we focus on users with a large number of tweets over the data collection period. There were 223 users with over 600 tweets and 115 users with over 1,000 tweets.

### Validating Model Assumptions

The coupled HMM framework developed in this paper is built on four main assumptions, two of which lead to the HMM formulation and two that couple the influence of the neighbors to the HMM. Notwithstanding the fact that these assumptions are based on a rational model of user behavior, the first two assumptions have been well-studied and justified in the literature [7, 18–20]. We now provide some empirical results to justify the latter two assumptions.

For this, we start with two typical users (denoted as User-I and User-II) whose activity over the thirty-day period consists of: i) 807 tweets, 260 mentions, and 16,935 tweets from his social network of 62 friends, and ii) 1,914 tweets, 1,108 mentions, and 10,281 tweets from his social network of 92 friends. We also consider an extreme case of a highly active user (denoted as User-III) whose activity over the thirty-day period consists of 2,387 tweets, 2,872 mentions, and 58,810 tweets from his social network of 206 friends. Users-I and II do not appear to be popular public figures, whereas User-III is a popular journalist, advocate on many political issues, and an activist.

Assumption 3 hypothesizes that each user’s activity switches from a baseline dynamics to an elevated state of dynamics depending on the magnitude of  $\tau$ . To test this assumption, we study the data from User-III for  $n = 1000$ . With this data, we use the generalized Baum-Welch algorithm to learn model parameters (as a function of  $\tau$ ) for a coupled HMM with the number of mentions as the influence structure. Fig. 3 plots the learned transition probabilities for User-III as a function of  $\tau$ . From this figure, we see that both  $p_0$  and  $p_1$  start off around approximately the same value (and similarly for  $q_0$  and  $q_1$ ). However, as  $\tau$  increases, the transition probabilities stabilize at different values suggesting that there is indeed a baseline and an elevated state in user dynamics. Similar behavior is also seen with data from Users-I and II.

On the other hand, Assumption 4 hypothesizes that  $Z_i$  and  $Z_{i-1}$  are conditionally independent given  $Q_{i-1}$ . To test this assumption, we use the model parameters learned with the generalized Baum-Welch algorithm in the generalized Viterbi algorithm to estimate the most likely state sequence corresponding to the observations. The conditional correlation coefficient between  $Z_i$  and  $Z_{i-1}$ , defined as,

$$\rho(Z_i, Z_{i-1}|Q_{i-1} = j) \triangleq \frac{\mathbb{E}\left[(Z_i - \mathbb{E}[Z_i|Q_{i-1} = j]) \cdot (Z_{i-1} - \mathbb{E}[Z_{i-1}|Q_{i-1} = j])\right]}{\sqrt{\mathbb{E}\left[(Z_i - \mathbb{E}[Z_i|Q_{i-1} = j])^2\right]} \cdot \sqrt{\mathbb{E}\left[(Z_{i-1} - \mathbb{E}[Z_{i-1}|Q_{i-1} = j])^2\right]}}, \quad j \in \{0, 1\}$$

is used to study conditional independence. Table 6 lists  $\rho(Z_i, Z_{i-1}|Q_{i-1} = j)$  and the  $p$ -value corresponding to this coefficient for Users-I to III with  $n = 800$ ,  $n = 1000$  and  $n = 1000$ , respectively. From this table, we see that the correlation coefficient in all the six cases studied has a small (absolute) value. A simple

explanation for this observation is that user aggregation significantly diminishes the correlation between  $Z_{i-1}$  and  $Z_i$ . Specifically, at a (standard) significance level of 5%, the null hypothesis of conditional independence between  $Z_i$  and  $Z_{i-1}$  cannot be rejected in five of the six cases studied indicating that Assumption 4 can be justified for many users.

### Model Fits For Users-I to III

We now study the following models for the activity profile of the three users: i) conventional two-state HMM, ii) coupled HMM with a binary influence structure that is set to 1 when there is a mention of the user and 0 otherwise, iii) coupled HMM with the number of such mentions as the influence structure, and iv) coupled HMM with the social network traffic of the friends of the user as the influence structure. Exponential and gamma densities are considered for the observations (inter-tweet duration). On the other hand, geometric, Poisson and shifted zeta densities are considered for the number of mentions, and a geometric density is considered for the total traffic. See Tables 1 and 2 for model details.

Tables 7-9 list the AIC scores for these three users with the different models as a function of the number of observations  $n$  for different choices:  $n = 100$ ,  $n = 250$ ,  $n = 500$ ,  $n = 750$ , or  $n = 1000$ . From these tables, the following conclusions can be made:

1. For all the three users, both a Poisson process model and a renewal process model are *significantly* sub-optimal for the observations as they implicitly assume a single state for the user's activity. This is in conformance with similar observations in [7]. A conventional HMM with two states overcomes this problem by assuming that the user switches between an *Active* and an *Inactive* state and thus provides a better baseline to compare the performance of the proposed modeling framework.
2. For all combinations of users,  $n$  and types of influence structure, a two-parameter gamma density for the observations results in a better fit than possible with an exponential model. This should not be entirely surprising since an exponential density is a special case of the gamma density (a gamma with  $k = 1$  and  $\lambda = \frac{1}{\rho}$  results in an exponential of rate  $\rho$ ). Similar observations have also been made in related recent work [14–17].

This conclusion is also reinforced by observing the Q-Q plots of the true inter-tweet durations (in the *Active* and *Inactive* states) relative to the inter-tweet duration values obtained from four models for User-III (see Figs. 4(a)-(d) and Figs. 4(e)-(h), respectively). The four competing models illustrated are: i) Model A — conventional HMM with exponential density, ii) Model B — conventional HMM

with gamma density, iii) Model C — coupled HMM with geometric influence structure and exponential density, and iv) Model D — coupled HMM with geometric influence structure and gamma density. The most probable state sequence vector corresponding to the observations of User-III is estimated with the (generalized) Viterbi algorithm. As can be seen from Fig. 4, Model D is the best fit from among these four models. Nevertheless, the discrepancies of some of the quantiles from the reference straight-line shows that even this model does not *completely* capture all the features in the observations, suggesting a direction for future work in this area.

3. We now explain how a coupled HMM works. State estimation with Models B and D is performed using the (generalized) Viterbi algorithm. In the  $n = 1000$  case, while the HMM declares 898 of the 1000 inter-tweet periods (corresponding to 21.06% of the total observation period for User-III) as *Active*, the coupled HMM declares only 845 periods (corresponding to 16.46% of the total observation period) as *Active*. In terms of discrepancies in state estimation between the two models, 53 inter-tweet periods declared as *Active* by the HMM are re-classified as *Inactive* by the coupled HMM, whereas all the *Inactive* states of the HMM are also classified as *Inactive* by the coupled HMM. Carefully studying  $\Delta_i$ ,  $Q_{i-1}$  and  $Z_i$  for these 53 periods that are re-classified, it can be seen that the coupled HMM declares a period as *Inactive* (independent of the nature of  $Q_{i-1}$  or  $Z_i$ ) provided that  $\Delta_i$  is large, or when  $\Delta_i$  is small and in addition,  $Z_i$  is also small and  $Q_{i-1} = 0$ . In other words, if the user is in the *Inactive* state and the influence structure does not suggest a switch to the *Active* state, a small inter-tweet period is treated as an *anomaly* rather than as an indicator of change to the *Active* state. Thus, in contrast to the HMM setting where the state estimate depends primarily on the magnitude of  $\Delta_i$ , the coupled HMM is less *trigger-happy* in the sense that it considers the magnitude of  $\Delta_i$  in the context of neighbors' activity before declaring a state as *Active* or *Inactive*.
4. In terms of general trends with AIC as the metric for model fitting, if  $n$  is small (say,  $n = 100$  to 250), a conventional HMM is competitive and comparable with (sometimes, even better than) a coupled HMM with more parameters. In addition to there not being sufficient data to learn a complicated model, this trend conforms with the popular intuition of the Occam's razor that simplistic models shall suffice for observations of small length.
5. Social network traffic (that includes replies, retweets, and modified tweets from all of a user's friends) typically overwhelms the number of mentions by at least an order of magnitude (see typical examples with Users-I and II above). Thus, social network traffic serves as a good influence structure to couple

a HMM when the number of mentions is too small to learn a sophisticated model reliably. This is typically the case when  $n$  is moderate (neither too small nor too large). For example, with Users-I and II, traffic leads to a better model fit for  $n = 500$  and  $n = 100$ , respectively.

6. However, as  $n$  increases, the more *directional* nature of a mention (relative to the traffic) means that mentions carry more “information” about the capacity of a user to respond/reply conditioned on seeing a certain type of tweet from his network (than the traffic). This is clear from the general trend of lower AIC scores with the number of mentions than with the social network traffic for large  $n$  values ( $n = 750$  or  $1000$ ).
7. For all combinations of users and  $n$ , the binary influence structure for the mentions results in a poorer fit than the number of mentions. This is because it is more efficient to capture the number of mentions with a one parameter model than to expend that parameter on a binary value. In other words, the loss in performance is due to the use of a hard decision metric (binary value), provided that the soft decision metric (the number of mentions) is captured accurately. While the three models (geometric, Poisson and shifted zeta) for the number of mentions result in comparable performance for Users-I and II, the geometric results in a superior fit for User-III. Thus, a geometric density can serve as a robust model choice for the number of mentions. Given that the shifted zeta density captures heavy-tails, the above trend also suggests that the number of mentions over an inter-tweet duration is not likely to be heavy-tailed.

### Performance Across Users

Given that the exponential observation density consistently under-performs in model fitting relative to a gamma density, we henceforth focus on the performance of Model D with Model B as the baseline. This performance gain is captured by the relative AIC gain metric, defined as,

$$\Delta\text{AIC} \triangleq \text{AIC}\Big|_{\text{Model B}} - \text{AIC}\Big|_{\text{Model D}}. \quad (6)$$

In Fig. 5 and Table 10,  $\Delta\text{AIC}$  values are presented for different  $n$  values for Users-I to III. As can be seen from this data, Model B performs better than Model D for  $n < 550$  for User-I and Model D gets better as  $n$  increases after that. For User-III, Model D is better than Model B for all  $n > 200$  and the performance gain improves with increasing  $n$  till  $n = 800$  and then slightly decreases after that. In general, the performance gain with Model D for the typical user is negligible for small values of  $n$  and this gain improves (in general) as  $n$  increases.

To study this aspect more carefully, we now consider a corpus of 100 users with different numbers of tweets and mentions over their periods of activity. For all the users studied, it is observed that a local optimum (to reasonable accuracy) is achieved by the generalized Baum-Welch algorithm within 20-30 iterations and independent of the model parameter initializations. Fig. 6(a)-(b) plots the histogram of  $\Delta\text{AIC}$  for the corpus of 100 users with  $n = 500$  and  $n = 1000$ , respectively. From Fig. 6, it can be seen that Model D significantly out-performs Model B for a large fraction of the users and this improvement gets better as  $n$  increases.

To understand this, recall that  $\exp\left(-\frac{\Delta\text{AIC}}{2}\right)$  is the likelihood that Model B minimizes the information loss relative to Model D. Thus, a  $\Delta\text{AIC}$  value larger than 4.61 and 9.21 leads to a relative likelihood of 10% and 1%, respectively. For the corpus studied here, Model D is 100 times as likely to minimize information loss for 25% of the users at  $n = 500$  and 72% of the users at  $n = 1000$ , respectively. With a more relaxed benchmark, Model D is ten times as likely to minimize information loss for 33% and 85% of the users at  $n = 500$  and 1000, respectively.

For predictive performance, analogous to  $\Delta\text{AIC}$  in (6), we define the relative SMAPE gain metric as

$$\Delta\text{SMAPE} \triangleq \text{SMAPE}\Big|_{\text{Model B}} - \text{SMAPE}\Big|_{\text{Model D}}.$$

Fig. 6(c) plots the histogram of  $\Delta\text{SMAPE}$  for  $n = 500$  and it can again be seen that Model D is better than Model B in terms of predictive power for a large fraction of users. Thus, a coupled HMM provides a better modeling paradigm for the activity of a large set of users in social networks.

### User Clustering

After learning the model parameters for each user, we now consider the similarity between different users as implied by those parameters. Since the dataset from [11] has only around 200 users with sufficient activity ( $n > 500$ ) to learn general probabilistic models where the coupled HMM parameters learned via the generalized Baum-Welch algorithm converge to local optima in the model parameter space, we focus on a corpus of 150 of these users. The mean number of tweets and mentions for this corpus is 975.40 and 629.17, respectively. The mean number of friends and followers for the corpus is 251.83 and 206.46, respectively.

With the number of mentions as the influence structure, we focus on two coupled HMM-specific parameters that capture the interaction dynamic between a user and his social network to perform user clustering in the model parameter space. The parameter  $p_1/p_0$  measures the propensity (likelihood) of a user to become active upon seeing a large number of mentions in his timeline relative to a lack of such mentions:

$$\frac{p_1}{p_0} = \frac{\text{P}(Q_i = 1 | Q_{i-1} = 0, Z_i > \tau)}{\text{P}(Q_i = 1 | Q_{i-1} = 0, Z_i \leq \tau)}.$$

On the other hand, the parameter  $\gamma_1/\gamma_0$  measures the propensity of the user’s network to respond with a mention upon seeing activity at the user relative to his inactivity:

$$\frac{\gamma_1}{\gamma_0} = \frac{\text{P}(Z_i > 0 | Q_{i-1} = 1)}{\text{P}(Z_i > 0 | Q_{i-1} = 0)}.$$

Three natural clusters can be identified in the model parameter space as a function of the  $\left\{ \frac{p_1}{p_0}, \frac{\gamma_1}{\gamma_0} \right\}$  values:

- 1) The baseline scenario where the users are *not* significantly influenced by their neighbors and *vice versa* corresponds to  $\frac{p_1}{p_0} = 1 = \frac{\gamma_1}{\gamma_0}$ . The users for whom  $\frac{p_1}{p_0} < 1$  are the “tails” in the model space corresponding to this baseline scenario.
- 2) On the other hand, a large value of  $\frac{p_1}{p_0}$  indicates that more mentions can induce a user to the *Active* state. Restated, a user can be induced to post at a higher frequency by an active social network.
- 3) Similarly, a large value of  $\frac{\gamma_1}{\gamma_0}$  indicates that a user’s social network can be induced to become active (with a larger number of mentions) by the user’s activity.

Motivated by this argument of three natural clusters, Fig. 7 clusters 150 users using the  $K$ -means algorithm for  $K = 3$ . The result of this clustering is that 92 users belong to Cluster 1 centered around  $\left( \frac{p_1}{p_0}, \frac{\gamma_1}{\gamma_0} \right) = (1, 1)$ . Of the remaining 58 users, 35 belong to Cluster 2 where  $\frac{p_1}{p_0} > 1.5$  and 23 belong to Cluster 3 where  $\frac{\gamma_1}{\gamma_0} > 1.5$ .

To paraphrase the above discussion, Cluster 1 is made of a majority of the users corresponding to the baseline scenario. On the other hand, Clusters 2 and 3 consist of users outside Cluster 1 and those who are either tightly knit to their network (or *vice versa*). These users are significantly affected by their social influence. Some of the typical attributes/qualities that can best describe users in Cluster 2 are: commentarial, activist, garrulous, argumentative, opinionated, etc. Illustrating these facets, a sample activity listing over a single session of a typical user in Cluster 2 is provided in Table 11. The session begins with a question of “yeh.watching” (apparently, a cricket match between Pakistan and England) by a friend of the user. This is followed by a conversation between friends and unsolicited commentaries/observations on the ongoing match by the user-of-interest. Such active commentary is typical of this user’s posting behavior. Another sample argument between two users in Cluster 2 is provided in Table 12. This argument is an exchange of political opinions with each user trying to convince the other about their respective positions. At the end of the argument, one of the users realizes and acknowledges that he has become more vocal on social media, yet also sensitive to other users’ positions.

On the other hand, the social network of users in Cluster 3 share similar attributes as users in Cluster 2 even though the users themselves are often more reluctant to follow suit. To illustrate this subtle difference in behaviors, a sample activity listing of a typical user in Cluster 3 is presented in Table 13. Here, we see that the user’s social network is strongly opinionated in response to a news story introduced by the user.

Despite introducing the story, the user himself is not sufficiently polarized/aggressive in his response on social media. Similarly, after introducing another story, the user blindly agrees with other users' positions and jokes on the matter. Thus, these examples illustrate how the coupled HMM paradigm introduced in this work captures broad features on user behavior despite not capturing the textual content in any detail.

## Conclusion

We have introduced a new class of coupled Hidden Markov Models to describe temporal patterns of user activity which incorporate the social effects of influence from the activity of a user's neighbors. While there have been many works on models for user activity in diverse social network settings, our work is the first to incorporate social network influence on a user's activity. We have shown that the proposed model results in better explanatory and predictive power over existing baseline models such as a renewal process-based model or an uncoupled HMM. User clustering in the model parameter space resulted in clusters with distinct interaction dynamics between users and their networks. Specifically, three clusters corresponding to: a baseline scenario of no influence of a user on his network (and *vice versa*), and two clusters with significant influence of a user on his network, and the network on the user, respectively are identified.

While our work has developed a social network-driven user activity model, it has only scratched the surface in this promising arena of research. It would be useful to pursue a more detailed study of different candidate models for the influence structures and the observations. It is also of interest in understanding which type of tweets/posts (mentions, replies, retweets, undirected tweets) or the total traffic carries more "information" in developing good models for explanation and prediction at the individual scale. It would also be of interest to develop hierarchical social influence-driven models for groups of users as well as better understand those facets of a user's social network that influence him the most. In particular, a careful study of other network influence structures (such as transfer entropy-weighted traffic [23]) that can capture the interaction dynamic between a user and his network is of importance.

Combining temporal activity patterns with unstructured information such as the topic or nature of discussion, textual content, etc., could result in much better predictive performance than temporal activity alone. Further, understanding the contribution of the temporal and the textual parts of such a composite model in prediction would also be useful in understanding the limits and capabilities of activity profile modeling at the individual scale.

## List of Abbreviations

1. HMM – Hidden Markov Model
2. Coupled HMM – Coupled Hidden Markov Model
3. Q-Q Plot – Quantile-Quantile Plot
4. AIC – Akaike Information Criterion
5. SMAPE – Symmetric Mean Absolute Percentage Error

## Competing Interests

The authors declare that they have no competing interests.

## Authors' Contributions

VR performed the numerical experiments. All the authors developed the model, designed the experiments, interpreted the results, wrote the paper, and approved the final manuscript.

## Acknowledgment

This work was supported by the Defense Advanced Research Projects Agency (DARPA) under grant # W911NF-12-1-0034 at the University of Southern California. The authors would like to thank Sofus Macskassy for providing the **Twitter** dataset that was used in illustrating the algorithms proposed in this work.

## References

1. Barabási AL: **The origin of bursts and heavy tails in human dynamics**. *Nature* 2005, **435**:207–211.
2. Goh KI, Barabási AL: **Burstiness and memory in complex systems**. *Europhysics Letters* 2008, **81**(48002).
3. Vázquez A, Oliveira JG, Dezső Z, Goh K, Kondor I, Barabási AL: **Modeling bursts and heavy tails in human dynamics**. *Physical Review E* 2006, **73**(3):036127.
4. Vázquez A, Racz B, Lukacs A, Barabási AL: **Impact of non-Poissonian activity patterns on spreading processes**. *Physical Review Letters* 2007, **98**(15):158702.
5. Malmgren RD, Stouffer DB, Motter AE, Amaral LAN: **A Poissonian explanation for heavy tails in e-mail communication**. *Proceedings of the National Academy of Sciences* 2008, **105**(47):18153–18158.
6. Malmgren RD, Stouffer DB, Campanharo A, Amaral LAN: **On universality in human correspondence activity**. *Science* 2009, **325**(5948):1696–1700.
7. Malmgren RD, Hofman JM, Amaral LAN, Watts DJ: **Characterizing individual communication patterns**. *In Proceedings of ACM SIGKDD Conference on Knowledge, Discovery, and Data Mining, Paris, France* 2009, :607–615.
8. Stehlé J, Barrat A, Bianconi G: **Dynamical and bursty interactions in social networks**. *Physical Review E* 2010, **81**(3):035101.

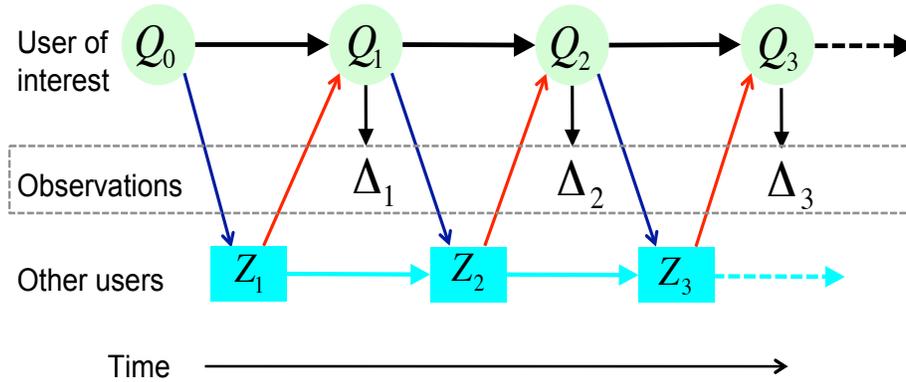
9. Rybski D, Buldyrev SV, Havlin S, Liljeros F, Makse HA: **Communication activity in a social network: relation between long-term correlations and inter-event clustering.** *Scientific Reports* 2012, **2**:560.
10. Jo HH, Karsai M, Kertész J, Kaski K: **Circadian pattern and burstiness in mobile phone communication.** *New Journal of Physics* 2012, **14**:013055.
11. Macskassy SA: **On the study of social interactions in Twitter.** In *Proceedings of International Conference on Weblogs and Social Media, Dublin, Ireland* 2012, :226–233.
12. Guo L, Chen S, Xiao Z, Tan E, Ding X, Zhang X: **Measurements, analysis and modeling of BitTorrent-like systems.** In *Proceedings of ACM SIGCOMM Internet Measurement Conference, New Orleans, LA* 2005, :35–48.
13. Leskovec J, Backstrom L, Kumar R, Tomkins A: **Microscopic evolution of social networks.** In *Proceedings of ACM SIGKDD Conference on Knowledge, Discovery, and Data Mining, Las Vegas, NV* 2008, :462–470.
14. Xiao Z, Guo L, Tracey J: **Understanding instant messaging traffic characteristics.** In *Proceedings of IEEE International Conference on Distributed Computing Systems, Toronto, Canada* 2007, :51.
15. Guo L, Tan E, Chen S, Xiao Z, Zhang X: **The stretched exponential distribution of Internet media access patterns.** In *Proceedings of ACM Symposium on Principles of Distributed Computing, Toronto, Canada* 2008, :283–294.
16. Guo L, Tan E, Chen S, Zhang X, Zhao Y: **Analyzing patterns of user content generation in online social networks.** In *Proceedings of ACM SIGKDD Conference on Knowledge, Discovery, and Data Mining, Paris, France* 2009, :369–378.
17. Jiang ZQ, Xie WJ, Li MX, Podobnik B, Zhou WX, Stanley HE: **Calling patterns in human communication dynamics.** *Proceedings of the National Academy of Sciences* 2013, **110**(5):1600–1605.
18. Wu Y, Zhou C, Xiao J, Kurths J, Schellnhuber HJ: **Evidence for a bimodal distribution in human communication.** *Proceedings of the National Academy of Sciences* 2010, **107**(44):18803–18808.
19. Hogg T, Lerman K: **Stochastic models of user-contributory web sites.** In *Proceedings of the International Conference on Weblogs and Social Media, San Jose, CA* 2009, :50–57.
20. Romero DM, Galuba W, Asur S, Huberman BA: **Influence and passivity in social media.** *Social Science Research Network Working Paper Series* 2010.
21. Rodriguez MG, Leskovec J, Krause A: **Inferring networks of diffusion and influence.** In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC* 2010, :1019–1028.
22. Zhao K, Bianconi G: **Social interactions model and adaptability of human behavior.** *Frontiers in Physiology* 2011, (2):101.
23. Steeg GV, Galstyan A: **Information transfer in social media.** In *Proceedings of World Wide Web Conference, Lyon, France* 2012, :509–518.
24. Tan C, Tang J, Sun J, Lin Q, Wang F: **Social action tracking via noise tolerant time-varying factor graphs.** In *Proceedings of ACM SIGKDD Conference on Knowledge, Discovery, and Data Mining, Washington, DC* 2010, :1049–1058.
25. Tan C, Lee L, Tang J, Jiang L, Zhou M, Li P: **User-level sentiment analysis incorporating social networks.** In *Proceedings of ACM SIGKDD Conference on Knowledge, Discovery, and Data Mining, San Diego, CA* 2011, :1397–1405.
26. Trusov M, Bodapati AV, Bucklin RE: **Determining influential users in Internet social networks.** *Journal of Marketing Research* 2010, **XLVII**:643–658.
27. Perra N, Gonçalves B, Pastor-Satorras R, Vespignani A: **Activity driven modeling of time varying networks.** *Scientific Reports* 2012, **2**:469.
28. Rabiner LR: **A tutorial on hidden Markov models and selected applications in speech recognition.** *Proceedings of the IEEE* 1989, **77**(2):257–286.
29. Bilmes JA: **A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models.** Tech. rep., International Computer Science Institute, Berkeley, CA 1998.

30. Pearl J: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann 1988.
31. Murphy KP: **Dynamic Bayesian Networks: Representation, Inference and Learning**. Tech. rep., University of California, Berkeley 2002. [Ph.D. dissertation].
32. Poritz AB: **Linear predictive hidden Markov models and the speech signal**. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Paris, France* 1982, **7**:1291–1294.
33. Juang BH, Rabiner LR: **Mixture autoregressive hidden Markov models for speech signals**. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 1985, **ASSP-33**(6):1404–1413.
34. Kenny P, Lennig M, Mermelstein P: **A linear predictive HMM for vector-valued observations with applications to speech recognition**. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 1990, **38**(2):220–225.
35. Cacciatore T, Nowlan SJ: **Mixtures of controllers for jump linear and non-linear plants**. In *Proceedings of Neural Information Processing Systems, Denver, CO* 1993, **7**:719–726.
36. Bengio Y, Fransconi P: **Input-output HMM's for sequence processing**. *IEEE Transactions on Neural Networks* 1996, **7**(5):1231–1249.
37. Ghahramani Z, Jordan MI: **Factorial hidden Markov models**. *Machine Learning* 1997, **29**:1–31.
38. Brand M, Oliver N, Pentland A: **Coupled hidden Markov models for complex action recognition**. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico* 1997, :994–999.
39. Pavlovic V: **Dynamic Bayesian networks for information fusion with applications to human-computer interfaces**. Tech. rep., University of Illinois, Urbana-Champaign, IL 1999. [Ph.D. dissertation].
40. Kwon J, Murphy KP: **Modeling freeway traffic with coupled HMMs**. Tech. rep., University of California, Berkeley, CA 2000.
41. Chu S, Huang T: **Bimodal speech recognition using coupled hidden Markov models**. In *Proceedings of IEEE International Conference on Spoken Language Processing, Beijing, China* 2000, **2**:747–750.
42. Chu S, Huang T: **Audio-visual speech modeling using coupled hidden Markov models**. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL* 2002, :2009–2012.
43. Zhong S, Ghosh J: **HMMs and coupled HMMs for multi-channel EEG classification**. In *Proceedings of International Joint Conference on Neural Networks, Honolulu, HI* 2002, :1154–1159.
44. Dong W, Pentland A, Heller KA: **Graph-coupled HMMs for modeling the spread of infection**. In *Proceedings of Conference on Uncertainty in Artificial Intelligence, Catalina Is., CA* 2012, :227–236.
45. Karsai M, Kaski K, Barabási AL, Kertész J: **Universal features of correlated bursty behaviour**. *Scientific Reports* 2012, **2**:397.
46. Dempster AP, Laird NM, Rubin DB: **Maximum-likelihood from incomplete data via the EM algorithm**. *Journal of Royal Statistical Society, Part B* 1977, **39**:1–38.
47. Liporace LA: **Maximum likelihood estimation for multivariate observations of Markov sources**. *IEEE Transactions on Information Theory* 1982, **IT-28**(5):729–734.
48. Jordan MI, Jacobs RA: **Hierarchical mixtures of experts and the EM algorithm**. *Neural Computation* 1994, **6**(2):181–214.
49. Jebara T, Pentland A: **Maximum conditional likelihood via bound maximization and the CEM algorithm**. In *Proceedings of Neural Information Processing Systems, Denver, CO* 1998, **11**:494–500.
50. Salojärvi J, Puolamäki K, Kaski S: **Expectation maximization algorithms for conditional likelihoods**. In *Proceedings of the International Conference on Machine Learning, Bonn, Germany* 2005, **22**:753–760.
51. Gnanadesikan R: *Methods for Statistical Data Analysis of Multivariate Observations*. Wiley-Interscience, 2nd edition 1997.

## Figures

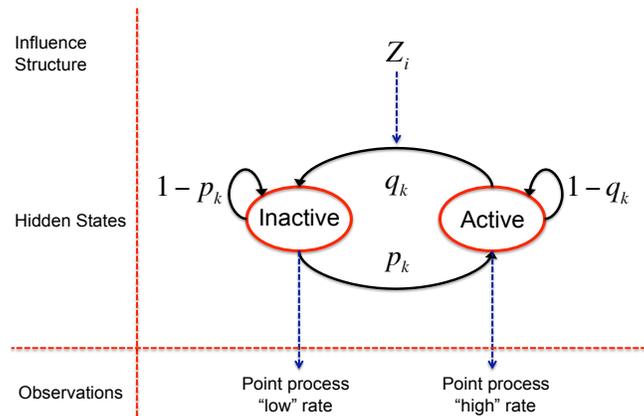
### Figure 1 - Coupled HMM

Coupled HMM framework for user activity.



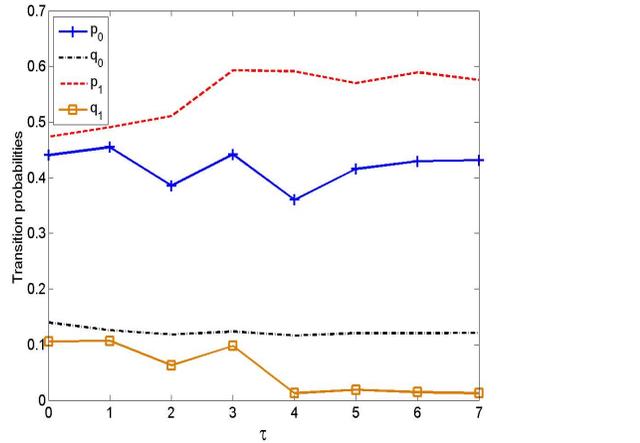
### Figure 2 - State transition

Pictorial illustration of state-transition evolution in the proposed model.



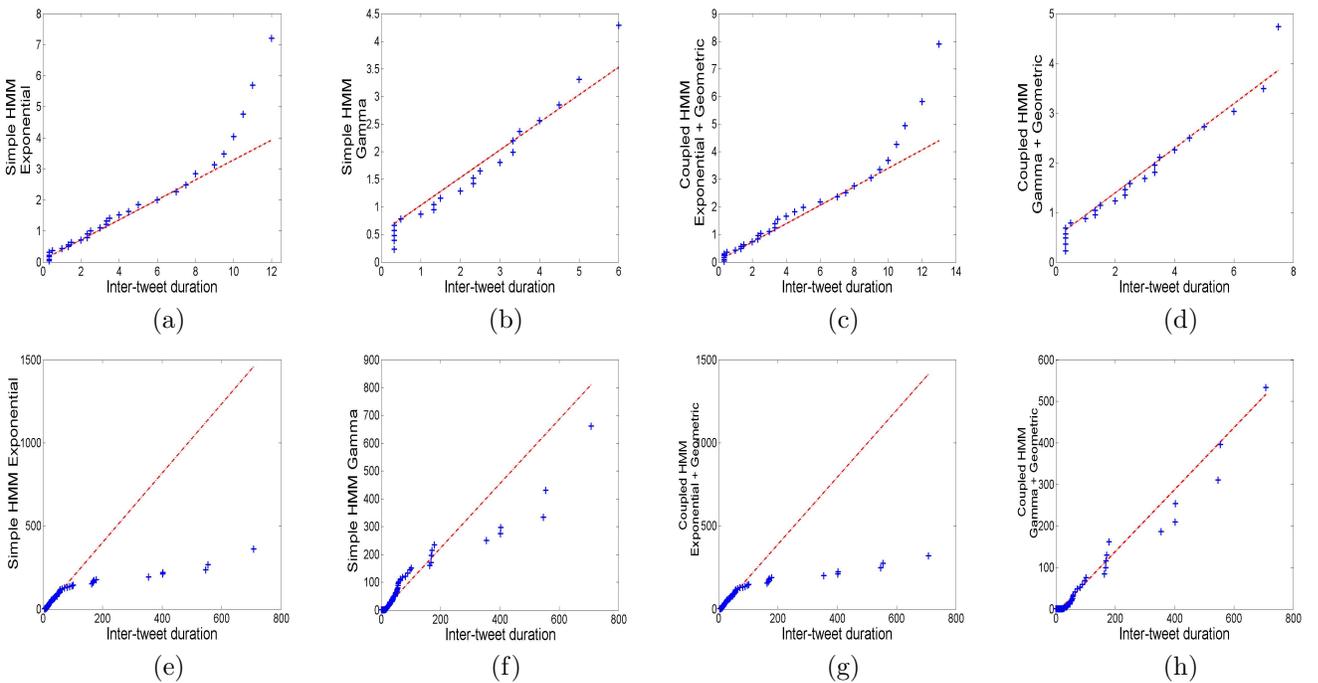
### Figure 3 - Model Validation

Learned transition probabilities as a function of  $\tau$  for User-III with  $n = 1000$ .



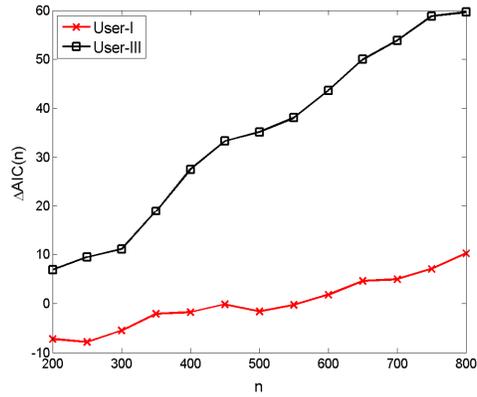
### Figure 4 - Q-Q Plots

Q-Q plots of inter-tweet duration in (a)-(d) *Active* and (e)-(h) *Inactive* states for User-III under four different models.



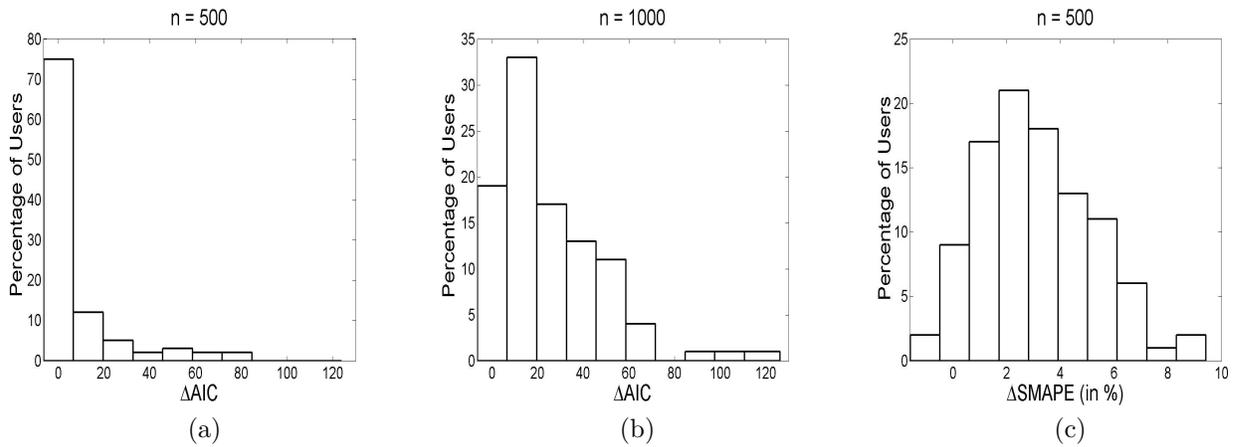
**Figure 5 - Performance Improvement**

AIC improvement for Model D relative to Model B.



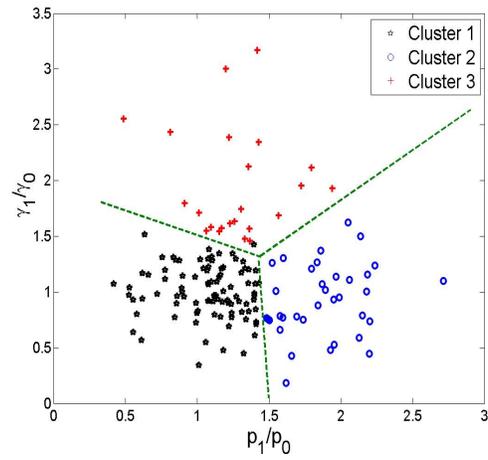
**Figure 6 - Histograms**

Histogram of AIC gain with coupled HMM for a corpus of 100 users with (a)  $n = 500$  and (b)  $n = 1000$  observations. (c) Histogram of SMAPE gain with coupled HMM for  $n = 500$ .



### Figure 7 - User Clustering

Clustering 150 users into three clusters based on learned model parameter values,  $p_1/p_0$  and  $\gamma_1/\gamma_0$ .



## Tables

**Table 1 - Observation Models**

Models for observation given state  $Q_i = j, j \in \{0, 1\}$ .

Name	Density function ( $f_j(\Delta_i)$ )	Baum-Welch parameter estimate
Exponential	$\rho_j \cdot \exp(-\rho_j \cdot \Delta_i)$	$\hat{\rho}_j = \frac{\sum_{i=1}^{N-1} (\xi_i(j,0) + \xi_i(j,1))}{\sum_{i=1}^{N-1} \Delta_i \cdot (\xi_i(j,0) + \xi_i(j,1))}$
Gamma	$\frac{1}{\Gamma(k_j)\lambda_j^{k_j}} \cdot \Delta_i^{k_j-1} \exp\left(-\frac{\Delta_i}{\lambda_j}\right)$ where $\Gamma(x)$ is the Gamma function	$\hat{k}_j$ and $\hat{\lambda}_j$ solve: $k_j \lambda_j = \frac{\sum_{i=1}^{N-1} \Delta_i \cdot (\xi_i(j,0) + \xi_i(j,1))}{\sum_{i=1}^{N-1} (\xi_i(j,0) + \xi_i(j,1))}$ $\log(\lambda_j) + \Psi(k_j) = \frac{\sum_{i=1}^{N-1} \log(\Delta_i) \cdot (\xi_i(j,0) + \xi_i(j,1))}{\sum_{i=1}^{N-1} (\xi_i(j,0) + \xi_i(j,1))}$ where $\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is the di-gamma function

**Table 2 - Influence Structure Models**

Models for influence structure given state  $Q_{i-1} = j, j \in \{0, 1\}$ .

Name	Density function ( $g_j(Z_i)$ )	Baum-Welch parameter estimate
Binary mentions	$P(Z_i = k   Q_{i-1} = j) = \begin{cases} 1 - \gamma_j & \text{if } k = 0 \\ \gamma_j & \text{if } k = 1 \end{cases}$	$\tilde{\gamma}_j = \frac{\sum_{i=1}^N \sum_{i: Z_i=1} (\tilde{\xi}_{i-1}(j,0) + \tilde{\xi}_{i-1}(j,1))}{\sum_{i=1}^N (\tilde{\xi}_{i-1}(j,0) + \tilde{\xi}_{i-1}(j,1))}$
No. of mentions		
Geometric	$P(Z_i = k   Q_{i-1} = j) = (1 - \gamma_j) \cdot \gamma_j^k, k \geq 0$	$\tilde{\gamma}_j = \frac{\sum_{i=1}^N Z_i \cdot (\tilde{\xi}_{i-1}(j,0) + \tilde{\xi}_{i-1}(j,1))}{\sum_{i=1}^N (Z_i + 1) \cdot (\tilde{\xi}_{i-1}(j,0) + \tilde{\xi}_{i-1}(j,1))}$
Poisson	$P(Z_i = k   Q_{i-1} = j) = \frac{\gamma_j^k \cdot \exp(-\gamma_j)}{\Gamma(k+1)}, k \geq 0$	$\tilde{\gamma}_j = \frac{\sum_{i=1}^N Z_i \cdot (\tilde{\xi}_{i-1}(j,0) + \tilde{\xi}_{i-1}(j,1))}{\sum_{i=1}^N (\tilde{\xi}_{i-1}(j,0) + \tilde{\xi}_{i-1}(j,1))}$
Shifted zeta	$P(Z_i = k   Q_{i-1} = j) = \frac{(1+k)^{-\gamma_j}}{\zeta(\gamma_j)}, k \geq 0$ where $\zeta(x)$ is the Riemann-zeta function	$\tilde{\gamma}_j$ solves: $\frac{-\zeta'(\gamma_j)}{\zeta(\gamma_j)} = \frac{\sum_{i=1}^N \log(1+Z_i) \cdot (\tilde{\xi}_{i-1}(j,0) + \tilde{\xi}_{i-1}(j,1))}{\sum_{i=1}^N (\tilde{\xi}_{i-1}(j,0) + \tilde{\xi}_{i-1}(j,1))}$

**Table 3 - Conditional Dependency Assumptions**

Conditional dependencies of the involved variables in different HMM architectures.

First-Order AR-HMM [32-34]	IO-HMM [36]	Coupled Factorial HMM [37]	Coupled HMM [38, 43]	Proposed Model
$\{Q_i, \Delta_{i-1}\} \rightarrow \Delta_i$	$\{Z_i, Q_i\} \rightarrow \Delta_i$	$\{Z_i, Q_i\} \rightarrow \Delta_i$	$Q_i \rightarrow \Delta_i$	$Q_i \rightarrow \Delta_i$
-	$Z_{i-1} \rightarrow Z_i$	$Z_{i-1} \rightarrow Z_i$	$\{Q_{i-1}, Z_{i-1}\} \rightarrow Z_i$	$Q_{i-1} \rightarrow Z_i$
$Q_{i-1} \rightarrow Q_i$	$\{Q_{i-1}, Z_i\} \rightarrow Q_i$	$\{Q_{i-1}, Z_{i-1}\} \rightarrow Q_i$	$\{Q_{i-1}, Z_{i-1}\} \rightarrow Q_i$	$\{Q_{i-1}, Z_i\} \rightarrow Q_i$

**Table 4 - Number of Parameters**

Number of parameters describing different models with  $k$  states,  $\ell$  network influence structure levels and an  $m$  parameter observation density in each state. Number of parameters for  $k = \ell = 2$  and  $m = 1$  or  $m = 2$  are also provided.

No. of parameters Model ( $\downarrow$ )	Trans. prob. matrix	Obs. density	Influence structure	Total	$m$	
					1	2
Conventional HMM	$k(k-1)$	$km$	-	$k(k+m-1)$	4	6
First-Order AR-HMM	$k(k-1)$	$k(m+1)$	-	$k(k+m)$	6	8
IO-HMM	$\ell k(k-1)$	$\ell km$	$\ell(\ell-1)$	$\ell(k(k+m-1)+\ell-1)$	10	14
Coupled Fact. HMM	$\ell k(k-1)$	$\ell km$	$\ell(\ell-1)$	$\ell(k(k+m-1)+\ell-1)$	10	14
Coupled HMM [38]	$(\ell+k)(k-1)$	$km$	$(\ell+k)(\ell-1)$	$(\ell+k)(\ell+k-2)+km$	10	12
Coupled HMM [43]	$2(\ell+k)(k-1)$	$km$	$2(\ell+k)(\ell-1)$	$2(\ell+k)(\ell+k-2)+km$	18	20
Proposed Model	$\ell k(k-1)$	$km$	$k(\ell-1)$	$k(\ell k+m-1)$	8	10

**Table 5 - Algorithm Derivation**

Steps in generalized Baum-Welch and Viterbi algorithms. Here,  $a, b$  and  $c$  are state variable values and  $i$  is the time-index.

**Steps in generalized Baum-Welch algorithm:**

$$\tilde{\xi}_i(a, b) = \frac{\tilde{\alpha}_i(a) \text{P}(Z_{i+1}|Q_i = a) \text{P}(Q_{i+1} = b|Q_i = a, Z_{i+1}) \text{P}(\Delta_{i+1}|Q_{i+1} = b) \tilde{\beta}_{i+1}(b)}{\sum_{a,b} \tilde{\alpha}_i(a) \text{P}(Z_{i+1}|Q_i = a) \text{P}(Q_{i+1} = b|Q_i = a, Z_{i+1}) \text{P}(\Delta_{i+1}|Q_{i+1} = b) \tilde{\beta}_{i+1}(b)}, \text{ where} \quad (7)$$

$$\tilde{\alpha}_i(a) \triangleq \text{P}(\Delta_1^i, Z_1^i, Q_i = a), \quad 2 \leq i \leq N - 1 \quad (8)$$

$$= \left[ \sum_b \tilde{\alpha}_{i-1}(b) \text{P}(Z_i|Q_{i-1} = b) \text{P}(Q_i = a|Q_{i-1} = b, Z_i) \right] \text{P}(\Delta_i|Q_i = a) \quad (9)$$

$$\tilde{\alpha}_1(a) = \left[ \sum_b \pi_b \text{P}(Z_1|Q_0 = b) \text{P}(Q_1 = a|Q_0 = b, Z_1) \right] \text{P}(\Delta_1|Q_1 = a) \quad (10)$$

$$\tilde{\beta}_i(b) \triangleq \text{P}(\Delta_{i+1}^N, Z_{i+1}^N|Q_i = b), \quad 2 \leq i \leq N - 1 \quad (11)$$

$$= \text{P}(Z_{i+1}|Q_i = b) \left[ \sum_c \text{P}(Q_{i+1} = c|Q_i = b, Y_{i+1}) \text{P}(\Delta_{i+1}|Q_{i+1} = c) \tilde{\beta}_{i+1}(c) \right] \quad (12)$$

$$\tilde{\beta}_N(b) = 1. \quad (13)$$

**Steps in generalized Viterbi algorithm:**

$$\delta_1(a) \triangleq \text{P}(Q_1 = a, \Delta_1, Z_1|\boldsymbol{\lambda}) \quad (14)$$

$$= \left[ \sum_b \text{P}(Q_0 = b) \text{P}(Z_1|Q_0 = b) \text{P}(Q_1 = a|Q_0 = b, Z_1) \right] \cdot \text{P}(\Delta_1|Q_1 = a) \quad (15)$$

$$\phi_1(a) = 0 \quad (16)$$

$$\delta_{i+1}(a) \triangleq \max_{Q_1^i} \left[ \text{P}(Q_1^i, Q_{i+1} = a, \Delta_1^{i+1}, Z_1^{i+1}|\boldsymbol{\lambda}) \right], \quad 1 \leq i \leq N - 1 \quad (17)$$

$$= \max_b \left[ \delta_i(b) \text{P}(Z_{i+1}|Q_i = b) \text{P}(Q_{i+1} = a|Q_i = b, Z_{i+1}) \right] \text{P}(\Delta_{i+1}|Q_{i+1} = a) \quad (18)$$

$$\phi_{i+1}(a) = \arg \max_b \left[ \delta_i(b) \text{P}(Z_{i+1}|Q_i = b) \text{P}(Q_{i+1} = a|Q_i = b, Z_{i+1}) \right]. \quad (19)$$

**Table 6 - Conditional Correlation Coefficient**

Conditional correlation coefficient  $\rho(Z_i, Z_{i-1}|Q_{i-1})$  for Users-I to III in *Active* and *Inactive* states.

User ( $\downarrow$ )	$Q_{i-1} = 1$			$Q_{i-1} = 0$		
	No. of samples	Corr. coef.	$p$ -value	No. of samples	Corr. coef.	$p$ -value
User-I	660	0.1127	0.0037	138	-0.0244	0.7762
User-II	909	0.0590	0.0753	89	-0.0540	0.6155
User-III	844	-0.0459	0.1832	154	0.1306	0.1065

**Table 7 - AIC Comparison for User-I**

AIC scores for a typical user with different model fits. The bolded figures correspond to the model with the best fit.

Model ( $\downarrow$ )	AIC			
	$n = 100$	$n = 250$	$n = 500$	$n = 750$
Observation density: Exponential				
Poisson process model	1017.71	2711.88	5157.41	7556.58
Conventional two-state HMM	835.42	2155.51	3852.15	5630.03
Coupled HMM, Binary mention	844.19	2169.88	3867.18	5658.23
Coupled HMM, No. mentions (Geometric)	840.65	2163.20	3853.10	5621.40
Coupled HMM, No. mentions (Poisson)	840.96	2163.54	3853.57	5621.90
Coupled HMM, No. mentions (Shifted zeta)	841.93	2163.39	3853.83	5623.03
Coupled HMM, Social network traffic (Geometric)	836.22	2156.11	3951.16	5789.35
Observation density: Gamma				
Renewal process model	879.02	2284.01	4214.99	6230.80
Conventional two-state HMM	<b>830.66</b>	<b>2134.12</b>	3806.88	5578.28
Coupled HMM, Binary mention	838.81	2148.02	3816.54	5595.53
Coupled HMM, No. mentions (Geometric)	836.36	2141.94	3808.53	<b>5571.18</b>
Coupled HMM, No. mentions (Poisson)	837.42	2142.66	3810.11	5572.50
Coupled HMM, No. mentions (Shifted zeta)	837.76	2142.13	3809.29	5584.69
Coupled HMM, Social network traffic (Geometric)	834.81	2149.96	<b>3797.65</b>	5583.40

**Table 8 - AIC Comparison for User-II**

AIC scores for a typical user with different model fits. The bolded figures correspond to the model with the best fit.

Model ( $\downarrow$ )	AIC		
	$n = 100$	$n = 500$	$n = 1000$
Observation density: Exponential			
Poisson process model	698.48	3444.09	6837.57
Conventional two-state HMM	504.20	2373.26	4343.94
Coupled HMM, Binary mention	517.35	2383.19	4354.92
Coupled HMM, No. mentions (Geometric)	504.62	2364.04	4333.98
Coupled HMM, No. mentions (Poisson)	504.42	2363.96	4333.78
Coupled HMM, No. mentions (Shifted zeta)	504.76	2363.94	4333.80
Coupled HMM, Social network traffic (Geometric)	489.87	2391.80	4406.64
Observation density: Gamma			
Renewal process model	613.98	2960.90	5691.82
Conventional two-state HMM	500.68	2334.20	4261.41
Coupled HMM, Binary mention	511.86	2339.44	4265.43
Coupled HMM, No. mentions (Geometric)	501.72	2320.02	4257.15
Coupled HMM, No. mentions (Poisson)	501.39	<b>2319.73</b>	<b>4256.61</b>
Coupled HMM, No. mentions (Shifted zeta)	501.92	2320.04	4256.93
Coupled HMM, Social network traffic (Geometric)	<b>487.70</b>	2336.06	4283.44

**Table 9 - AIC Comparison for User-III**

AIC scores for an extreme case of a highly active user with different model fits. The bolded figures correspond to the model with the best fit.

Model ( $\downarrow$ )	AIC		
	$n = 100$	$n = 500$	$n = 1000$
Observation density: Exponential			
Poisson process model	653.73	3141.36	6250.35
Conventional two-state HMM	479.31	2359.48	4611.80
Coupled HMM, Binary mention	495.63	2391.71	4677.30
Coupled HMM, No. mentions (Geometric)	478.69	2337.30	4552.67
Coupled HMM, No. mentions (Poisson)	483.11	2339.65	4682.30
Coupled HMM, No. mentions (Shifted zeta)	490.09	2370.70	4753.87
Coupled HMM, Social network traffic (Geometric)	499.81	2402.47	4663.76
Observation density: Gamma			
Renewal process model	585.78	2830.55	5594.71
Conventional two-state HMM	476.39	2290.53	4440.57
Coupled HMM, Binary mention	485.79	2318.99	4494.15
Coupled HMM, No. mentions (Geometric)	<b>473.56</b>	<b>2262.10</b>	<b>4394.30</b>
Coupled HMM, No. mentions (Poisson)	489.48	2262.22	4521.20
Coupled HMM, No. mentions (Shifted zeta)	484.44	2300.41	4469.11
Coupled HMM, Social network traffic (Geometric)	485.73	2313.83	4458.81

**Table 10 - AIC Improvement**

$\Delta$ AIC for Users-I to III with different  $n$  values.

User ( $\downarrow$ )	Small $n$		Moderate $n$		Large $n$	
User-I	-5.70	( $n = 100$ )	-1.65	( $n = 500$ )	7.10	( $n = 750$ )
User-II	-1.04	( $n = 100$ )	14.18	( $n = 500$ )	4.26	( $n = 1000$ )
User-III	2.83	( $n = 100$ )	28.43	( $n = 500$ )	46.27	( $n = 1000$ )

**Table 11 - Sample Activity Listing**

Sample activity of a typical user in Cluster 2.

Posted by	Intended for	Text content
@Friend	@Cluster-2-user	yeh.watching
@Cluster-2-user		RT @Non-Friend: In the U.S., you can text "FLOOD" to 27722 to donate \$10 to the #Pakistan Relief Fund. [URL] #helppakistan
@Cluster-2-user	@Friend	:D
@Cluster-2-user		Is it mandatory for Saeed Ajmal to bowl on short ball in every over #fail #PakCricket
@Friend	@Cluster-2-user	i think afridi should go for gull
@Cluster-2-user		Believe me when I say that I predicted that shoaib will get yardy in this over :D #PakCricket
@Cluster-2-user		That was a khoooni yorker by Umer #Gull Waqar ki yaad aa gai #PakCricket
@Friend		RT @Cluster-2-user: 300 not out by Boom Boom [URL] @Friend1 @Friend2 ... @Friend8
@Cluster-2-user		Baba dam darood aka Muhammad Yousaf has taken an un believable catch #PakCricket
@Friend	@Cluster-2-user	;-)
@Friend		RT @Cluster-2-user: #blog post 300 not out by Boom Boom [URL] #PakCricket #BoomBoom #Afridi
@Friend	@Cluster-2-user	now what? another scandal against #pakcricket #Pakistan for winning the match
@Cluster-2-user	@Friend	i think the scandal is going to be damaging bat of Swan on the yorker attempt b Shabby #PakCricket
@Cluster-2-user		RT @Friend: There's always second chances... it just depends on how hard you fight for it !!
@Cluster-2-user		RT @Friend: dedicated to men in Green hope this streak continues and we we the coming one as well.... [URL]
@Cluster-2-user		Hay Jazba Janoon to Himmat na haar... the formula behind the success of cricket team #PakCricket

**Table 12 - Sample Argument**

Sample argument between two typical users in Cluster 2.

Posted by	Intended for	Text content
@Cluster-2-user-2	@Cluster-2-user-1	quick question...is @salmantaseer one of the bad guys?? i thought the PPP was the lesser of the evils over there.. #justwondering
@Cluster-2-user-1	@Cluster-2-user-2	PPP lesser of the evils?! Hahahaha everyone working under Zardari is as evil as Bieber can ever dream of being.
@Cluster-2-user-2	@Cluster-2-user-1	aww..but sherry rehman is so sweet..and i love the rehman malik hairdo..and SMQ is so totally suave..btw who are the good guys??
@Cluster-2-user-1	@Cluster-2-user-2	wow man! You're kidding, right?
@Cluster-2-user-2	@Cluster-2-user-1	man!!why do i get the feeling that i just said like a totally not cool thing..
@Cluster-2-user-1	@Cluster-2-user-2	Err yes. Totally uncool :P
@Cluster-2-user-2	@Cluster-2-user-1	so no good guys huh??and btw yes i do think @fbhutto is totally wannabeish... :)
@Cluster-2-user-1	@Cluster-2-user-2	Yeah, I don't like her either. I like Imran Khan better.
@Cluster-2-user-2	@Cluster-2-user-1	i liked him till i saw a report ver his old 92 world cup winning team members said they never liked him..
@Cluster-2-user-2	@Cluster-2-user-1	if he culdnt get his own team members behind him then.....
@Cluster-2-user-1	@Cluster-2-user-2	Well, can't say anything about that. But he's a good person, a better one in politics at least.
@Cluster-2-user-2	@Cluster-2-user-1	well..sub continental politics na...everyone has a murky past or present...just depends on how deep one digs..no offence intended
@Cluster-2-user-1	@Cluster-2-user-2	None taken. But look at his hospital. Despite a few controversies, I can tell you firsthand that it's amazing. Good man.
@Cluster-2-user-1		#becauseoftwitter I've become more vocal in terms of venting.
@Cluster-2-user-1		#becauseoftwitter I've become more sensitive. I lose a follower and I go berserk.

**Table 13 - Sample Activity Listing**

Sample activity of a typical user in Cluster 3.

Posted by	Intended for	Text content
@Cluster-3-user		My maid told today; somebody threw one day old daughter wrapped in a polythene bag in Filth Drum outside a house. #humanitarian
@Cluster-3-user		I'll write on this very issue; daughter, poverty or an illegitimate child was she? why our social system allows us to take these plunges?
@Friend-1	@Cluster-3-user	did the baby survived? what has went wrong with people, they are becoming so heartless. :(
@Cluster-3-user	@Friend-1	haan she survived and somebody has adopted her.. only there.. but it is heart breaking..
@Friend-1	@Cluster-3-user	oh Thanks God. Yes it is really heart breaking.
@Friend-2	@Cluster-3-user	alive ???? =O
@Cluster-3-user	@Friend-2	yeah alive.. an issueless couple has adopted her
@Friend-2	@Cluster-3-user	Thank God! .... thts literally inhuman act! I mean the adopting parents should atleast get back to her parents and hang them!
@Cluster-3-user	@Friend-2	where would they find them? to hang
@Cluster-3-user		Half of Pakistanis say match fixing allegations against cricketers untrue [URL] #Cricket #Pakistan
@Friend-3	@Cluster-3-user	which means half of Pakistan believes the allegations of #Cricket #MatchFixing against the #Pakistan team to be true #Honesty #ICC
@Cluster-3-user	@Friend-3	ha ha ha yeah.. agreed
@Friend-4	@Cluster-3-user	The proper way of match fixing is thru umpires, didnt give collingwood out he made a big partnership & gave well set akmal out
@Cluster-3-user	@Friend-4	true.. collingwood has fixed match.. LOL