Flu Gone Viral: Syndromic Surveillance of Flu on Twitter using Temporal Topic Models

Liangzhe Chen*, K. S. M. Tozammel Hossain*, Patrick Butler, Naren Ramakrishnan, B. Aditya Prakash
Department of Computer Science, Virginia Tech, VA, USA
Email: {liangzhe, tozammel, pabutler, naren, badityap}@cs.vt.edu

***Abstract*—Surveillance of epidemic outbreaks and spread from social media is an important tool for governments and public health authorities. Machine learning techniques for nowcasting the flu have made significant inroads into correlating social media trends to case counts and prevalence of epidemics in a population. There is a disconnect between data-driven methods for forecasting flu incidence and epidemiological models that adopt a state based understanding of transitions, that can lead to sub-optimal predictions. Furthermore, models for epidemiological activity and social activity like on Twitter predict different shapes and have important differences. We propose a temporal topic model to capture hidden states of a user from his tweets and aggregate states in a geographical region for better estimation of trends. We show that our approach helps fill the gap between phenomenological methods for disease surveillance and epidemiological models. We validate this approach by modeling the flu using Twitter in multiple countries of South America. We demonstrate that our model can consistently outperform plain vocabulary assessment in flu case-count predictions, and at the same time get better flu-peak predictions than competitors. We also show that our fine-grained modeling can reconcile some contrasting behaviors between epidemiological and social models.**

I. Introduction

Online web searches and social media such as Twitter and Facebook have emerged as surrogate data sources for monitoring and forecasting the rise of public health epidemics. The celebrated example of such surrogate sources is arguably Google Flu Trends where user query volume for a handcrafted vocabulary of keywords is harnessed to yield estimates of flu case counts. Such surrogates thus provide an easy-to-observe, indirect, approach to understanding population-level events.

The recent research has brought intense scrutiny on Google Flu Trends, often negative. Lazer et al. [17] provide many reasons for Google Flu Trend's lackluster performance. Some of these reasons are institutional (e.g., a cloud of secrecy about which keywords are used in the model, affecting reproducibility and verification); some are operational (e.g., lack of periodic re-training); others could be indicative of more systemic problems, e.g., that the vocabulary for tracking might evolve over time, or that greater care is needed to distinguish which aspects of search query volume should be used in modeling. These problems are not unique to Google Flu Trends, they can resurface with other surveillance strategies.

Our work is motivated by such considerations and we aim to better bridge the gap between syndromic surveillance strategies and contagion-based epidemiological modeling such as SI, SIR, and SEIS [12]. In particular, while models of social activity have been inspired by epidemiological research, recent work [20], [26], [23] has shown that there are key aspects along which they differ from biological contagions. Specifically, evidence from [20], [9] shows that the activity profile (or the number new people using a hashtag/keyword) shows a power-law drop—in contrast standard epidemiological models exhibit an exponential drop [12]. Also, there is some evidence that hashtags of different topics show an exposure curve which is not monotonic, resembling a complex contagion [23].

We show that we can reconcile the apparently contrasting behaviors with a finer-grained modeling of biological phases as inferred from tweets. For example, sample tweets "Down with flu. Not going to school." and "Recovered from flu after 5 day, now going to the beach" denote different states of the users (also see Figure I(a)). We argue that correcting for which epidemiological state a user belongs, the social and biological activity time-series are actually similar. Hashtags and keywords merge users belonging to different epidemiological phases. We separate these states by using a temporal topic model. In addition, thanks to the finer-grained modeling, our approach gets better predictions of the incidence of flu-cases than direct keyword counting and also sometimes gets better predictions of flu-peaks than sophisticated methods like Google Flu Trends.

Our contributions are:

1) We propose a temporal topic model (HFSTM) for inferring hidden biological states for users, and an EM-based learning algorithm (HFSTM-FIT) for modeling the hidden epidemiological state of a user.
2) We show via extensive experiments using tweets from South America that our learner indeed learns meaningful word distributions and state transitions. Further, our method can better forecast the flu-trend as well as flu-peaks.
3) Finally, we show how once corrected for the state information using our learnt model, the social contagion activity profile fits better with standard epidemiological models.

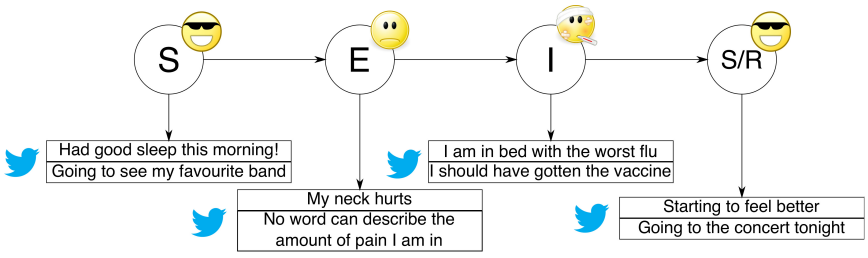
Our work can be seen as a stepping stone to better understanding of contagions that occur in both biological and social spheres.

II. Related Work

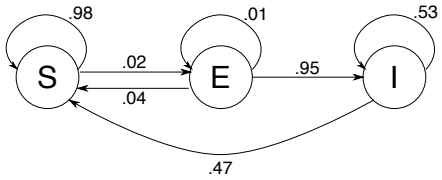
The most related work comes from three areas, we discuss them respectively in this section.

Epidemiology: In the epidemiological domain, various compartmental models (which explicitly model states of each

*Both authors contributed equally to the work.



(a) Toy example.



(b) State transition *learned* by our model HFSTM.

Fig. 1. Comparison between expected state transition and the state transition learned by our model. Figure (a) is a toy example showing possible user states and a tweet associated with each state. Figure (b) shows the state transition probability learned by HFSTM (see Sec. III).

user) are employed to study the characteristics of flu diffusion [12]. Some of the best known examples of such models are SI [14], SIR [3], and SEIS [19], which are regularly used to model true flu case counts. Recent works [20], [26] show that the social activity profiles do not exactly follow these models, and propose several other variants. Note that different epidemiological models are used for different diseases, in this paper we focus our work on flu since it is very common disease.

Social Media: The study of topic and word trends in social media has become an important predictor for real world events. These trends are much easier and faster to get from social media than from traditional methods (e.g. reliable CDC case counts typically have lags of more than a month). For disease prediction and forecasting, especially for flu, various methods have been proposed for large-scale [10] and small-scale predictions [8]. Furthermore, there are prediction methods that are solely based on Twitter [18]. Sadelik et al. [24] studied the impact of interactions to personal health. Lamb et al. [15] discriminate tweets that express awareness of the flu from those with actual infections, and train a classifier by which a user can tell if the author of a tweet is really infected. While their work is single-tweet-based, ours takes the tweet history into account. Achrekar et al. [1], and Lampos et al. [16] fit a flu trend by analysing tweets via various methods including keyword analysis, and compare their flu trend fitting with CDC results. These methods are very coarse-grained—they do not provide understanding on how the health state of a user changes over time, while we link the change of tweet pattern with standard epidemiological models.

Topic Models: We use a variation of topic models for our purposes. LDA (Latent Dirichlet Allocation) models are very popular for topic-modeling and many variations have been proposed. For modeling health related topics Paul et al. proposed the Ailment Topic Aspect Model (ATAM+) [22] to capture various ailments from a corpus of tweets. This model does not consider the temporal information of the text messages (as we do in this paper). A variant of LDA is temporal topic models which can be categorized into two groups: Markovian and non-Markov. Wang et al. [25] propose a non-Markov continuous time model for topic trends which can not be used to predict the user states. The Markovian methods [11], [2] only capture transition of topics within a document or a message, they do not capture state transition of users *across* tweets. There are two other variants of LDA [4], [13] studying the evolution of topic distributions over time, while our model studies the transition between a set of topic distributions which does not evolve over time. Moreover, their models do not capture the topic changes between consecutive messages of a user. Another recent related work is by Yang et al. [27] who combine keyword distributions with a shortest path algorithm to find out a monotonically increasing stage progression of an event sequence. In our problem, flu states changing are not monotonic, and we learn the transition probabilities, which their method does not.

TABLE I. SYMBOLS USED FOR HFSTM.

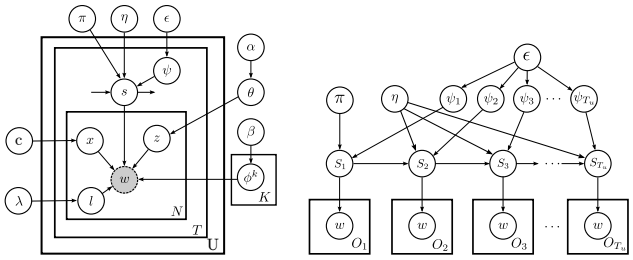
Sym.	Meaning	Sym.	Meaning
S	Flu state	S_t	Flu state for the t-th tweet
ψ	State switching variable	ϵ	Hyper-parameter for ψ
π	Initial state distribution	η	Transition probability matrix
x	Binary switcher between flu and non-flu words	l	Binary background switching variable
λ	Hyper-parameter for l	c	Hyper-parameter for x
θ	Topic distribution	ϕ	Topic-Word distribution
α	Hyper-parameter for topics	N_t	Number of words in t-th tweet
β	Dirichlet parameter for word distribution	T_u	Number of tweets for u-th user
w	Word variable	z	Non-flu related topic
w_{tn}	n-th word in t-th tweet	K	Number of states

III. MODEL FORMULATION

We formulate our model in this section. Our hypothesis is that a tweet stream generated by a user can be used to capture the underlying health condition of that particular user. We assume that the health state (e.g., flu state) of a user remains the same within a tweet. In this study we use our model to capture the flu states of a user which are S (healthy), E (exposed), or I (infected). We base it on the classic flu-like Susceptible-Exposed-Infected-Susceptible SEIS epidemiological model, which models the different states of a person throughout the lifecycle of the infection. We propose a Hidden Flu-State from Tweet Model HFSTM for modeling states from user's tweets.

A. HFSTM

A tweet is a collection of words and a tweet stream is a collection of tweets. The number of tweets varies across users and the number of words in a tweet varies within and across users. We denote the t-th tweet of a user by $O_t = \langle w_{t1}, w_{t2}, \ldots, w_{tN_t} \rangle$ where w_{tn} denotes the n-th word in the tweet and N_t denotes the total number of words in the tweet. Let $\mathcal{O}_u = \langle O_1, O_2, \ldots, O_{T_u} \rangle$ be the tweet stream generated by a user u and $\mathcal{S}_u = \langle S_1, S_2, \ldots, S_{T_u} \rangle$ be the underlying state of the stream $\mathcal{O}_u$. Here T_u denotes the length of the stream of a user u and $S_t \in \{S, E, I\}$. Let $\mathcal{O} = \langle \mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_U \rangle$



(a) HFSTM (b) State transition

Fig. 2. (a) Plate notation for HFSTM: The variable S captures the hidden state of the user in which the user generated this tweet. The LDA-like topic variable z capture non-flu related words. (b) HFSTM state variables expanded: Each message O_t is associated with a state S_t, which remains same for flu-related words in O_t. Switching from one state to another is controlled by a binary switching variable ψ and the next state S_{t+1} from the current state S_t is drawn using transition probabilities η.

Algorithm 1 Generator($\lambda, c, \eta, \pi, \alpha, \beta, \epsilon$)

Input: A set of parameters.
Output: Topics and flu state of each user.

1. Set the background switching binomial λ
2. Choose $\phi \sim$ Dir(β) for the non-flu topics, flu states, and background distribution
3. Choose initial state $s_1 \sim$ Mult(π)
4. Draw each row of η using Dir(α) ▷ Trans. matrix
5. Draw $\theta \sim$ Dir(α)
6. **for** each tweet $1 \leq t \leq T_u$ **do**
7. **if** not the 1st tweet in the corpus **then**
8. Draw $\psi_t \sim$ Ber(ϵ)
9. **if** $\psi_t = 0$ **then**
10. $S_t \leftarrow S_{t-1}$
11. **else**
12. $S_t \leftarrow$ Mult($\eta_{S_{t-1}}$)
13. **for** Each word w_i, $1 \leq i \leq N_t$ **do**
14. Draw $l_i \in \{0, 1\} \sim$ Ber(λ) ▷ Background switcher.
15. **if** $l_i = 0$ **then**
16. Draw $w_i \sim$ Mult(ϕ^B) ▷ Background distribution.
17. **else**
18. Draw $x_i \in \{0, 1\} \sim$ Ber(c)
19. **if** $x_i = 0$ **then**
20. Draw $z_i \sim$ Mult(θ)
21. Draw $w_i \sim$ Mult(ϕ^{z_i}) ▷ Non-flu related distribution.
22. **else**
23. Draw $w_i \sim$ Mult(ϕ^{s_t}) ▷ Flu related distribution.

be the collection of tweets for U users, from which we aim to learn the parameters of our model. We use K to denote the number of states that S_t can take (see Table I for notations).

Our method—Hidden Flu-State from Tweet model HFSTM—is a probabilistic graphical model which captures the tweet structure of a flu-related tweet. It is a temporal topic model for predicting the state sequence of a user given $\mathcal{O}_u$ and is illustrated in Fig. 2(a). An expansion of the plate notation for the same is illustrated in Fig. 2(b). In this model each word w for $O_t \in \mathcal{O}_u$ is generated when the user is in a particular flu state (S_t) or the user talks about a non-flu related topic (z_i). For example, in the message "I have caught the flu. Feeling feverish. Not going to school" the words 'flu', 'feverish', 'caught' are generated because the user is in the "infected" state and the words 'going' and 'school' are generated by non-flu related topics. Sometimes a word can be generated due to noise which is also accounted for in our model.

A generative process for the model is shown in Alg. 1. A binary variable l determines whether or not a word is generated from a background distribution. The binary variable x determines whether the current word is generated from non-flu related topics or flu-state distributions. The value of l and x are generated from Bernoulli distributions parameterized by λ and c. The non-flu related topics follow the LDA like mechanism [5]. The state for the first tweet is drawn from the initial distribution denoted by π. We assume that the states of the subsequent tweets are generated due to a state transition or by copying from the previous state which is determined by a binary switching variable ψ with prior parameter ϵ. The state S_t (for $2 \leq t \leq T_u$) of the subsequent tweets are drawn from transition matrix η and previous state S_{t-1} with probability ϵ or copied from the previous state S_{t-1} with probability $1 - \epsilon$. Once the state of a tweet is determined, a word is generated from a word distribution defined by that state.

Let $O_t = (w_1, \ldots, w_N)$ be the words that are generated when a user is in a particular state. The likelihood of the words generated by a user in that state is given below.

$$p(\mathcal{O}_u) = \sum_{S_t} p(\mathcal{O}_u, S_t) = \sum_{S_t} p(O_1 \ldots, O_T, S_t)$$
$$= \sum_{S_t} \sum_{S_{t-1}} p(O_t|S_t)p(S_t|S_{t-1})p(O_{t-1}, S_{t-1}) \quad (1)$$

Andrews et al [2] show that this likelihood function is intractable. In our model the unknown parameters that we want to learn are ϕ,c,λ,ϵ,η,π. The posterior distributions over these unknown variables are also intractable since the posterior distributuions depend on the likelihood function. We develop an EM-based algorithm HFSTM-FIT to estimate the parameters $H = (\epsilon, \pi, \eta, \phi, \lambda, c)$ of our model. We omit the equations here due to the lack of space.

IV. Experiments

A. Experimental Set-up

First we describe our set-up in more detail. Our algorithms were implemented in Python. [1]

1) Choosing Vocabulary: To ensure that the most important words (directly flu-related words like 'flu', 'cold', 'congestion', etc.) are included in our vocabulary, we first build a flu-related keyword list. Chakraborty et al. [7] construct a flu-keyword list, by first manually setting a seed set, then using two methods (pseudo-query and correlation analysis, see their paper for more details) to expand this seed set, and then finally pruning it to a 114 words keyword list. Note that this keyword list can be updated automatically if the flu vocabulary evolves. For our experiments, we include the same 114 keywords from Chakraborty et al. [7] first. We then include 116 words selected by our in-house experts, which are not directly related to flu, but may implicitly imply the state of a user, such as 'hopeless', 'bed', 'die', 'sad', etc. We use this mixture of automatically and manually generated (a total of 230) words[1], including a generic block-word which we map all other words to, as the vocabulary for HFSTM.

2) Datasets: We collected tweets generated from 15 countries in South America for the period Dec, 2012—Jan, 2014 using Datasift's Twitter collection service[2], which pre-processes the data and detects the geo-location for tweets.

[1]Code and vocabulary can be found here: http://people.cs.vt.edu/liangzhe/code.html

[2]http://datasift.com/

We create a training dataset *TrainData*, using the tweets from Jun 20, 2013 to Aug 06, 2013, which contains a peak of infections. We created two evaluation sets: *TestPeriod-1*, using tweets from Dec 01, 2012 to Jul 08, 2013, which contains the rising part of a flu infection peak; *TestPeriod-2*, from Nov 10, 2013 to Jan 26, 2014, which is from a different flu season. For creating training data we perform keyword and phrase checking (from our vocabulary) to identify a set of users who have potentially tweeted a flu-related tweet. We then fetch their tweet streams from Twitter API for the training period. We then use the Datasift service to preprocessing these tweets (stemming, lemmatization, etc.), and get our final training dataset of roughly 34,000 tweets.

We collected data from The Pan American Health Organization (PAHO [21]) for the ground-truth reference dataset for flu case counts (trends). PAHO plays the same role in South America as CDC does in the USA . Note that PAHO gives only per-week counts.

B. Word distritution for each flu-state

In short, our model learns meaningful topic word distribution for the flu states. See Figure 3–it shows a word cloud for each flu-state (we renormalized each word distribution after removing the generic block-word) learned by HFSTM. The most frequent words in each state matches well with the S, E and I states in epidemiology. As shown in the figure, the S state has normal words, the E state starts to gather words which indicate an exposure or approaching to the disease ('pain', 'throat'), while the I state gets many typical flu-related words ('flu', 'fever').

C. State transition

We show the state transition diagram learned by our model in Figure I(b). The initial state probability learned is $[0.98, 0.02, 0.00]$, with high probability that a tweet starts at state S, 0.02 probability it starts at state E, and almost zero probability it starts at state I. When there's a transition occurring, a tweet in S state tends to stay in S state, a tweet in E state is very likely to enter I state, while a tweet in I state either stays infected or recovers and goes back to state S. All these observations match closely with the standard epidemiological SEIS model and intuition.

We also investigate the most-likely state sequence for each user learned by our model. Using the probabilities learned by our model, we take a sequence of tweets from one user, and use MLE to estimate the state each tweet is in. Table II shows one example of these transitions (we show the translated English version here using Google Translate). As we can see, our model is powerful enough to learn the Exposed state, before the user is infectious. This also shows the accuracy of our transition probabilities between the flu states.

D. Fitting flu trend

Additionally, to test the predictive capability of our models, we design a flu-case count prediction task on our test datasets, after training on *TrainData*. We compare three models: (A) the baseline model, which uses classical linear regression techniques and word counts to predict case count numbers; (B) our models HFSTM; and (C) GFT (Google Flu Trend). In all three cases we use the same LASSO based linear regression model to predict the number of cases of influenza like illnesses recorded by PAHO (the ground-truth). We predict per-weekly values as both PAHO and GFT give counts only on a weekly basis.

Date	Tweet Message	State
29 Jul	I hate pork chops - . -	S
29 Jul	I just want to leave my house to eat what I like my	S
29 Jul	I'm dying of sleep , headache and sore throat but I will because I have mathematical	E
29 Jul	That itv program brainwashed my mom , now I want to take juice or eat cereal	S
29 Jul	Everything would be perfect if I hurt your throat	E
30 Jul	I'm sure I have a fever because I hear weird sounds	I
30 Jul	I will survive because I am macabre empire	I
30 Jul	I want to go to the doctor - . -	I
30 Jul	Natural orange juice for the sick	I
30 Jul	spicy ham tkm	I

TABLE II. EXAMPLE STATE SEQUENCE FOR A USER AS LEARNT FROM OUR MODEL FROM REAL-WORLD TWEETS (TRANSLATED TO ENGLISH USING GOOGLE TRANSLATE).
We used HFSTM to classify tweets to different states. As we can see, our model can capture the difference between different states and also the state transitions.

The baseline model uses a set of features created from the counts of 114 flu related words. We count the number of occurrance of these words in the testing data, these word counts were then collated into a single feature vector defined as the number of tweets containing a single word per week. We then regressed this set of counts to the PAHO case counts for each week.

Our models improve upon the baseline model by incorporating the state of the user when a word was tweeted. In this way we capture the context of a word/tweet as implied by our HFSTM models. For our models, the feature vector is created from a count of the top 20 words from each state, appended to the word of each state, such that $(cold, S)$ is counted differently from $(cold, I)$.

For GFT, we directly collect data from the Google Flu Trends website[3], and then apply the same regression as used in other methods to predict the number of infection cases. Note that as GFT is a state-of-the-art production system with highly optimized proprietary vocabulary lists, we do not expect to beat it consistently, yet as we describe later, we note some interesting results.

Fig. 4(a) shows the aggregated cases for *TestPeriod-1*, and Fig. 4(b) shows the smae cases for *TestPeriod-2*. We make several observations. Firstly, it is clear from the figures that HFSTM outperforms the baseline method (of keyword counting) in both cases—demonstrating that the state knowledge is important and our model is carefully learning that information correctly (the RMSE value difference between HFSTM and the baseline for the 2 plots are about $[250, 70]$ respectively). Secondly, we also see that the predictions from our model

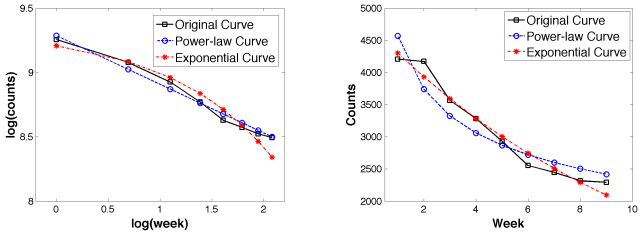
[3]http://www.google.org/flutrends

(a) S state (b) E state (c) I state

Fig. 3. The translated word cloud for the most probable words in the S, E and I state-topic distributions as learnt by HFSTM on *TrainData*. Words are originally learned and inferred in Spanish, we then translate the result using google translate for the ease of understanding. The size of the word is proportional to its probability in the corresponding topic distribution. Our model is able to tease out the differences in the word distributions between them.

are comparable qualitatively to the state-of-the-art GFT predictions, even though our method was just implemented as a research prototype without sophisticated optimizations. In fact, for Figures 4(b), our model HFSTM even *outperforms* GFT (with an RMSE difference of about 37). Significantly, in both cases, GFT clearly overestimates the peak which our method does not (this is an important issue with GFT which was also documented and observed in context of another US flu season as well [6]). These results show that including the epidemiological state information of users via our model can potentially benefit the prediction of infection cases dramatically.

E. Bridging the Social and the Epidemiological



(a) Total Keyword activity (log-log) (b) Key. Act. in learnt I state (lin-lin)

Fig. 5. Finer grained models help bridge the gap between social and epidemiological activity models. (a) Power law describes keyword activity better (in *log-log* axes to show the difference). (b) Exponential function explains well the falling part of the curves for keyword activity (note the *linear* axes).

Finally, as mentioned before, another key contribution of our model is to try to bridge the gap between epidemiological models and social activity models. An important recent observation [20], [26] was that the fall-part of any social activity profile is power-law—in contrast to standard epidemiological models like SEIR/SIR which give an exponential drop-off. How can they be reconciled? Next we show that accounting for the differences in the epidemiological state as learnt by our model, these activity profiles look the same i.e. they drop-off exponentially as expected from standard epidemiological models.

To test our hypothesis, we chose commonly occuring flu-keywords—enfermo (sick), fiebre (fever), dolor (pain)—for the analysis. Firstly, we count the total occurences of these keywords in *TestPeriod-1*. For each keyword we identify the falling part of its activity-curve. We then fit each curve with power law and exponential function. As expected from [20], Fig. 5(a) shows that the power-law function provides a much better fit of the falling part of the curve compare to the exponential function (RMSE scores for power law and exponential functions are ~ 320.31 and ~ 469.35 respectively).

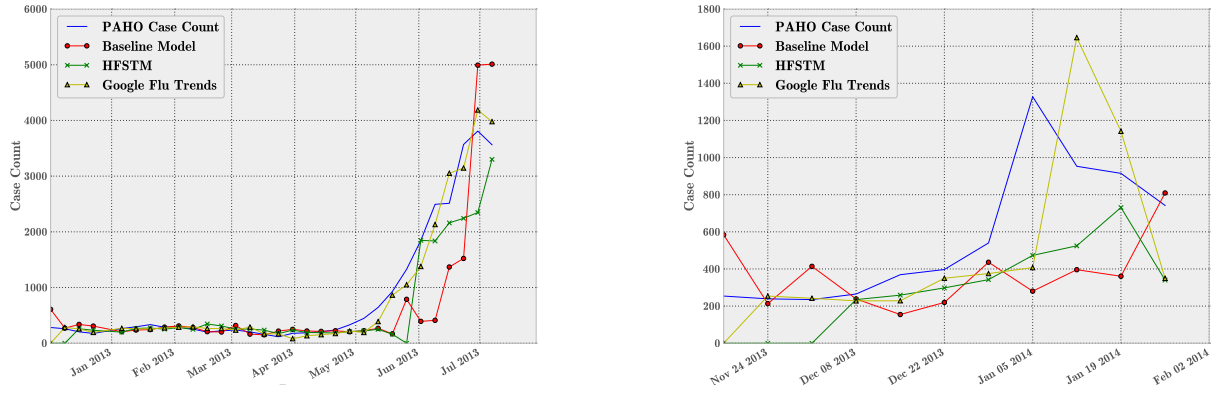
Secondly, to study the effect of our model on the activity profiles of these keywords: we count total occurrences of these keywords in the tweets which are tweeted *only by infected* users (i.e. by those users we learn as being in I). In contrast to the previous figure, we see that now exponential fit (RMSE score $\sim$ 147.48) is much better than a power law fit (RMSE score $\sim$ 275.50) (see Fig. 5(b))—matching what we would expect from an epidemiological model like SEIS. Thus this demonstrates that finer-grained modeling can explain differences between the biological activity and the social activity which is used as its proxy.

V. Discussion and Conclusion

Predicting the hidden state of a user from a sequence of tweets is highly challenging. Through extensive experiments, we showed how our method HFSTM can effectively model hidden states of a user and the associated transitions, and use it to improve flu-trend prediction, including avoiding recent errors discovered in methods like Google Flu Trends. We also showed how our model can reconcile seemingly different behaviors from social and epidemiological models, lending a state aware nature to data-driven models and simultaneously, letting simulation oriented models estimate their state transition matrices by maximizing data likelihood.

HFSTM uses unsupervised topic modeling, which means that the model itself does not discriminate between words. This would be a problem when the vocabulary contains many background words; where HFSTM may learn the unpredictable states behind background words because of the sparsity of the flu-related words. This is also the reason why we use a rather 'clean' vocabulary for HFSTM. One of the extensions we are currently working on is how to robustly learn meaningful flu-related states and topics even with an enlarged and noisier vocabulary.

Acknowledgements. This material is based upon work supported by the National Science Foundation under Grant No. IIS-1353346, by the Maryland Procurement Office under contract H98230-14-C-0127, by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D12PC000337, and by the VT College of Engineering. Any



(a) *TestPeriod-1* (b) *TestPeriod-2*

Fig. 4. Evaluation for the two test scenarios: (a) test on *TestPeriod-1*. (b) test on *TestPeriod-2*. Comparison of the week to week predictions against PAHO case counts using the three models: baseline model, HFSTM, and GFT (Google Flu Trend). Our model outperforms the baseline, and is comparable to GFT, beating it in case of (b). GFT overestimates the peak in both test periods.

opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the respective funding agencies.

REFERENCES

[1] H. Achrekar, A. Gandhe, R. Lazarus, S.-H. Yu, and B. Liu. Predicting Flu Trends using Twitter Data. In *IEEE Conference on Computer Communications Workshops*, pages 702–707. IEEE, 2011.

[2] M. Andrews and G. Vigliocco. The Hidden Markov Topic Model: A Probabilistic Model of Semantic Representation. *Topics in Cognitive Science*, 2(1):101–113, 2010.

[3] E. Beretta and Y. Takeuchi. Global Stability of an SIR Epidemic Model with Time Delays. *The Journal of mathematical biology*, 33(3):250–260, 1995.

[4] D. Blei and J. Lafferty. Dynamic Topic Models. In *In ICML*, pages 113–120, 2006.

[5] D. Blei, A. Ng, and M. Jordan. Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

[6] D. Butler. When Google got Flu Wrong. *Nature*, 494(7436):155–156, 2013.

[7] P. Chakraborty, P. Khadivi, B. Lewis, A. Mahendiran, J. Chen, P. Butler, E. Nsoesie, S. Mekaru, J. Brownstein, M. Marathe, and N. Ramakrishnan. Forecasting a Moving Target: Ensemble Models for ILI Case Count Predictions. In *SIAM International Conference on Data Mining*, 2014.

[8] N. Christakis and J. Fowler. Social Network Sensors for Early Detection of Contagious Outbreaks. *PLoS ONE*, (9), 09 2010.

[9] R. Crane and D. Sornette. Robust Dynamic Classes Revealed by Measuring the Response Function of a Social System. In *PNAS*, 2008.

[10] J. Ginsberg, M. Mohebbi, R. Patel, L. Brammer, M. Smolinski, and L. Brilliant. Detecting Influenza Epidemics using Search Engine Query Data. *Nature*, 457(7232):1012–1014, 2008.

[11] A. Gruber, M. Rosen-Zvi, and Y. Weiss. Hidden Topic Markov Models. *Artificial Intelligence and Statistics (AISTATS)*, 2007.

[12] H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42, 2000.

[13] L. Hong, D. Yin, J. Guo, and B. Davison. Tracking Trends: Incorporating Term Volume into Temporal Topic Models. In *the 17th ACM SIGKDD*, pages 484–492, 2011.

[14] J. Jacquez and C. Simon. The Stochastic SI Model with Recruitment and Deaths I. Comparison with the Closed SIS Model. *Mathematical Biosciences*, 117(1):77–125, 1993.

[15] A. Lamb, M. J. Paul, and M. Dredze. Separating fact from fear: Tracking flu infections on twitter. In *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2013.

[16] V. Lampos, T. De Bie, and N. Cristianini. Flu detector: Tracking epidemics on twitter. In *Proceedings of the 2010 European Conference on Machine Learning and Knowledge Discovery in Databases: Part III*, ECML PKDD'10, pages 599–602, 2010.

[17] D. Lazer, R. Kennedy, G. King, and A. Vespignani. The parable of google flu: Traps in big data analysis. *Science*, 343(6176):1203–1205, 2014.

[18] K. Lee, A. Agrawal, and A. Choudhary. Real-time disease surveillance using twitter data: Demonstration on flu and cancer. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*, pages 1474–1477. ACM, 2013.

[19] M. Li and J. Muldowney. Global stability for the SEIR model in epidemiology. *Mathematical Biosciences*, 125(2):155–164, 1995.

[20] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: model and implications. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '12, pages 6–14, 2012.

[21] PAHO. Epidemic disease database, pan american health organization. http://ais.paho.org/phip/viz/ed_flu.asp, Dec. 2012.

[22] M. Paul and M. Dredze. You Are What You Tweet: Analyzing Twitter for Public Health. In *Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011)*, 2011.

[23] D. M. Romero, B. Meeder, and J. Kleinberg. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 695–704, 2011.

[24] A. Sadilek, H. Kautz, and V. Silenzio. Predicting disease transmission from geo-tagged micro-blog data. In *AAAI Conference on Artificial Intelligence*, 2012.

[25] X. Wang and A. McCallum. Topics Over Time: a non-Markov Continuous-time Model of Topical Trends. In *the 12th ACM SIGKDD*, pages 424–433, 2006.

[26] J. Yang and J. Leskovec. Patterns of temporal variation in online media. In *WSDM*, pages 177–186, 2011.

[27] J. Yang, J. McAuley, J. Leskovec, P. LePendu, and N. Shah. Finding progression stages in time-evolving event sequences. In *the 23rd International Conference on World Wide Web*, pages 783–794, 2014.