The Attention Automaton: Sensing Collective User Interests in Social Network Communities

Suman Deb Roy, Gilad Lotan and Wenjun Zeng

***Abstract*—The vast quantity of information shared in social networks has brought us to an age of attention scarcity, where getting users to be attentive to a message is not a given. In fact, it has become the limiting factor in the consumption of information by end users. Understanding what captures the collective attention within a community of users in a social network is invaluable to many applications, such as product marketing, targeted advertising and social or political campaign organization. Several scholars have analyzed how information spreads in social networks under the constraint of attention. However, few papers provide a quantitative method to model and predict attention at every instant in the dynamic social web.**

In this paper, we propose the *Attention Automaton*, a probabilistic finite automata that can estimate the collective attention of some user community. Communities are based on geographical vicinity of users or having common interests (like followers of a given account) on Twitter. We identify two key factors that drive collective user attention: (1) the attention *volatility* of the community (frequency of change of trending topics), and (2) the selective categorical affinity of the user group towards certain trends. Our results, which are based on a 8-month dataset of Twitter trending topics across 111 geographic regions and audience trends of approximately 50 brands indicate that the proposed *Attention Automaton* can predict audience reception of impending trends based on categorical filters and inherent oscillations in user activity.

***Index Terms*— attention, automata, media, news, twitter, trending topics, trends, volatility**

I. Introduction

Social networks such as Facebook and Twitter emerged as a platform for connecting people who wanted to stay in touch, be heard, share information and voice opinions. With an increasing number of users, brands and highly visible celebrities joining these services, there came an inevitable explosion in the amount of content readily available to users. Social network data has dwarfed every other kind of traffic on the web, and truly ushered the age of Big Data.

The statistics on the amount of data generated via social network channels is astonishing. For example, there are 8500 Facebook likes per second. Twitter users send over 100,000 tweets per minute. YouTube has more than 1 billion unique users each month and approximately 100 hours of video are uploaded every minute. Almost 2500 Foursquare check-ins are performed every minute. Every 60 seconds on Facebook: 510 comments are posted, 293,000 statuses are updated, and 136,000 photos are uploaded. This hyperactivity is indicative that users and content creators have unrivalled freedom to publish as much as they want online [34].

As the threshold to publishing nears zero, getting users to be attentive turns into a limiting factor in our networked information ecosystem [1]. One cannot demand attention, or even expect it at a given point in time. It is a scarce commodity that must be earned [2]. Studying what users pay attention to is critical to many web applications, including product marketing, targeted advertising, social and political campaigns etc. Moreover, in a world where our attention is dissected in various ways every single day, it is fascinating to explore what can sustain user interest. Knowledge of where attention is given helps web intelligence algorithms to personalize user experience.

We can quantify user attention within social networks by looking at the level of interest that a node (user) dedicates in managing its interaction with another node or group of nodes within the observed social network [3]. The interaction can be captured in different activities, such as 'liking' a Facebook status update, 're-tweeting/RT' a tweet or posting in relation to some trending topic on Twitter [18]. For example, we can consider that when node X on Twitter retweets (RTs) a message M of another node Y, then X was attentive to Y or to the content of M. Similarly, if node X tweets about a topic that is trending, then we can claim that X was attentive towards that trend. By extension, when a group of users RT a certain tweet, they display *collective attention* [4]. This group of users could be geographically co-located, or followers of the same user or part of a networked community. Using these signals could be useful in analyzing information diffusion, attention drifts and network interdependency in between different subsections/communities of the network.

Taking one step forward, understanding the dynamics of collective attention is very useful, helping content producers and intermediaries better manage information flows under the constraint of human attention. It also brings clarity in judgment of what, when and why some trend becomes popular, which has great relevance to monetization of online content [5]. Social advertising utilizes a user and their community within social networked spaces, attempting to accurately target contextually relevant personalized ads [6]. 'Promoted Content' on Twitter is a good example of targeted social ads. In keyword advertising, advertisers need to predict

Suman Deb Roy is a Senior Data Scientist with Instapaper and Betaworks. Gilad Lotan is the Chief Scientist at Betaworks in New York City. Wenjun Zeng is a Professor at the Department of Computer Science, University of Missouri.

which keywords to buy [7]. This necessitates prior knowledge of the facets that capture audience attention. There are several other scenarios, including finding advertising leads (potential advertisers), which require collective attention modeling.

Here, let us first provide a motivating example regarding what the attention automaton has achieved in deployment. We demonstrate an app that was built to assist news editors. The app is a trend bot that detects trending news/stories using trend detection and attention estimation in Twitter (Fig. 1). It is currently being used in the social news website Digg.com, (www.digg.com) which has approximately 6 million monthly users. The tool analyzes 1.5 million twitter events per hour, covering 250 thousand web domains and generates a list of 5-10 links per 5 minutes for the editors to choose from. The basic functionality is driven by the Attention Automaton (proposed system described in Section IV), which spots trends within some user group that will persist for the longest duration. We have found that the persistence of the trend depends on the categorical affinity of the group to the trend category as well as the volatility, as explained later in Section IV. Once the trend is spotted, the bot selects a news story/ link for the suitable for the user group and recommends it. In doing so, the bot predicts which article will get most attention within the user group and allows news editors to select the optimal link to be featured for certain user groups.

The tool is extremely useful in the news production cycle, where trends can be quickly detected via Twitter and then matched to some user community interests via the Attention Automaton. This provides the user group with a story they will be most attentive towards, based on their categorical affinity and volatility. It also allows the news editors and producers to directly align their selection to user interests within network communities.

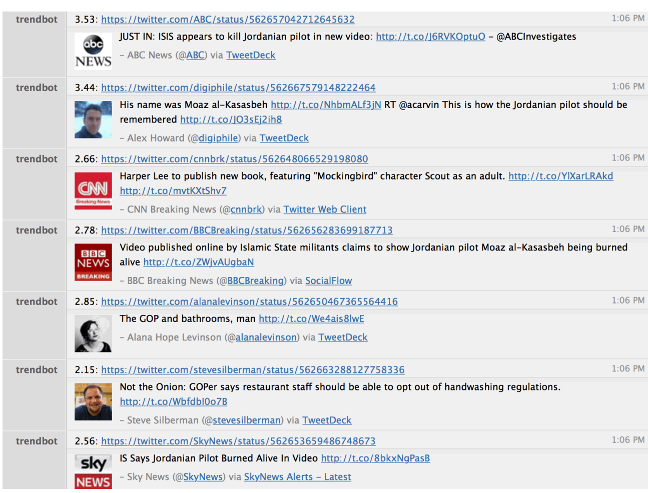

Fig. 1. The figure shows a trend bot built on top of the attention automaton. The bot uses the attention automaton to select which trend will persist within the audience trends in some user group/brand (e.g., CNN) and selects news stories/links based on those trends.

Previous research has attempted to capture the dynamics of popularity and information diffusion in social networks to get a sense of what receives user attention. Interesting findings from these works show that attention is the deciding factor in information spread [8], that there are specific categories which potentially receive more attention [4] and that these categories remain relatively consistent over long periods of time [3]. However, most of these works aim to understand individual user attention and miss the insights provided by the larger community [9]. The dynamics of a networked group of users reveals collective social intelligence - which is one of the primary traits in human interactions [10]. If we think of these phenomena as a computing model (e.g., in terms of finite state machines) that can represent the dynamics of collective attention, we can attempt to predict future collective behavior.

In this paper, we propose a probabilistic automaton that can capture the dynamics of collective attention among user groups on Twitter, who are either geographically co-located or co-followers of the same Twitter account. Every state in the automaton is a list of trending topics of the user group. Over time, the *Attention Automaton* transitions from one state to another, mirroring the changes in the trending topics list (TTL), as shown in Fig. 2. The action that causes the state transitions is a set of competing trends that are trying to break into the TTL at any given moment (impending trends). When a trend breaks into the TTL, it forces the automaton to jump to a new state, as the list (TTL) changes.

The probability of transition depends on two key factors: (1) the inherent attention shift tendency of the group, and (2) the selective categorical affinity of the group towards certain trends. The inherent attention shift tendency of the user group is modeled using a metric we call *Volatility*, which represents how often trending topics within the group change over time. We use a Levenshtein distance [11] based metric to formulate the volatility of the user group. The selective categorical affinity is obtained from the past history of topics that trended in the user group and their respective categories.

Our results reveal interesting information about the social network user communities:

- Two opposing forces drive collective attention: *volatility vs. categorical affinity of attention*.
- The collective attention of user groups over time on Twitter can be modeled as a probabilistic automaton. This automaton has predictive power over future states given a time series of impending trends.
- The *Attention Automaton* can capture Twitter community reactions to real world events.

The rest of this paper is organized as follows: In Section II we discuss related work and motivation. Section III contains the data description and insights from data. The proposed Attention Automaton model is described in

Section IV. In Section V, we provide model evaluation results. We conclude the paper in Section VI with future work.

II. Related Work and Scope

We present related work in the realm of attention in social networks and the motivation for this work.

A. Attention Economics

Human attention is the mental 'spotlight' on a stage full of information. Herbert Simon first laid out the idea that attention is a scarce commodity [12]. However, it was Davenport et al. who first indicated that attention precedes activity on the web [13]. Given the overload of information in cyber space, search engines and recommendation systems attempt to learn from our interactions (click through etc.) to identify and predict resources that users would be more attentive towards [14]. Moreover, advertising metrics of cost-per-click-through or cost-per-thousand-viewers is strongly dependent on a quantifiable measure for attention [15]. Understanding the attention of online communities can be very useful for advertising leads, targeted advertising and marketing [6]. There have been efforts to identify influence spreaders in complex networks and quantify attention on Wikipedia documents to stock market variations [38, 39].

B. Attention in Social Networks

Attention is captured by the behavior of social network nodes in the face of competing choices of interaction [16]. It has been found that attention is the primary barrier for social contagion and information propagation in online social networks [8]. Huberman et al. showed that the allocation of attention among a set of items in social news website Digg is log-normally distributed [17]. Lehmann et al. found that the evolution of hash tags popularity in Twitter follows discrete classes, indicating user groups are attentive to selected categories of information [4]. The balance of attention dedicated to these categories is a relatively stable property over time [3]. Researchers have also shown that a combination of social network structure and finite attention is a sufficient condition for emergence of dynamics of social networks [18]. Search query results can also be correlated with what has captured mass attention [37]. This makes attention modeling in social networks a vital prerequisite to predict future popularity and lifetime of trends.

C. Probabilistic Automaton

Finite state machines are fundamental to computer science. They are widely used as spelling checkers and Hidden Markov models [19]. A probabilistic automata is a state transition system, consisting of a series of states, actions that can cause transition between states, and a probability attached to each potential transition from one state to another [20]. The input to the automaton could be a time series of actions. Relationship between probabilistic automata, Hidden Markov models and learning algorithms is discussed in [21].

D. Twitter Trends

When a group of users within a defined geographic region increasingly post the same hashtag, word or phrase, it may be identified as a Twitter trend. At any given point in time, Twitter's trending topics list (TTL) includes a list of the top 10 trends (Fig. 1). The precise algorithm for determining what trends on Twitter is proprietary, but the basic thrust is that it's not only about volume, but always normalized compared to previous points in time [30]. Thus, a new topic that is discussed by Twitter users at rates higher than the usual has higher likelihood to become a Twitter trend [32]. Social information regarding breaking news, celebrity gossip and political campaigns often emerge as trending topics as well [22, 24, 33].

E. Scope of the paper

Since attention precedes online activity [13] it is pivotal to model attention of user communities in order to comprehend the fundamental differences in behavior between user groups, in other words, what makes them unique. Here we present the scope of our paper within network science and engineering.

Firstly, there are two key limitations in existing work in this domain: (1) Although social data mining reveals popularity and novelty of trends as a good indicator of attention patterns of users [17], it still does not help us quantify the collective attention shifts in communities or the categorical attention affinity that exists in users groups [34]. Most importantly, it gives us few indications as to whether collective attention is at all *computable* (in terms of a model of computation) and whether we can predict the likelihood of a future trend to receive sustained attention. (2) The dynamics of collective attention is substantially different from individual attention [9]. Our analysis shows that a collection of users bound together as followers of a given account or within close geographic proximity can play a big role in what becomes popular and receives attention. For a motivating real-world example, see the Appendix at the end (Fig. 12).

Our research strives to address these limitations. We build a probabilistic automaton, called the *Attention Automaton*, showing that attention states in user communities of Twitter are computable in terms of finite state machines. Moreover, unlike previous work, we focus on collective attention which is endorsed by the collective behavior of the inherent communities we are part of in the social network.

Finally, our paper explores new techniques of modeling social network properties. Section IV. B provides a novel way of network sampling and measurement of information diversity in Twitter communities. Section IV. C describes modeling of network dynamics in the face of competing information and trends. We provide a fast real-time technique of analyzing information shifts and attention-capturing topics without requiring expensive graph (topological) process.

As mentioned in the Introduction/ Section I, we developed a novel tool to assist journalists and news editors in quickly figuring out breaking news, fine-tuned by the community of readership (sports readers vs. politics readers). It analyzes 1.5 million twitter events per hour covering 250 thousand web domains based on the Attention Automaton. The experiments discussed in this research (Section V) are conducted on, to the best of our knowledge, the largest Twitter trending topics data ever used in social network trends research.

III. Data

Our research is based on two datasets containing Twitter trend data. Since our method is substantially data-driven, we first provide a description of the data. The first data set includes Twitter trends based on geographical locations for approximately 8 months. To our best knowledge, this is the largest dataset of trends used in any research on Twitter trends. The second data set contains trends for audiences who are subscribers/followers of some Twitter account for approximately 3 months. Again, to the best of our knowledge, this is both one of the first and the largest dataset of audience trends ever used for research. Both trends are represented as time series signals over their duration. We briefly describe the individual datasets below.

A. *Geographical Trends (GT)*

When a group of users on Twitter increasingly RT a message or tweet about some topic, then it is captured as a trend. We wrote a script that probes Twitter every 5 minutes and logs the TTL (Fig. 2) provided by Twitter for 111 geographical locations worldwide (for details see: https://blog.twitter.com/2012/see-trends-for-100-more-cities). Although Twitter uses a proprietary algorithm to calculate 'trends' from tweets, it is best known to be a combination of tf-idf on a bag-of-words model of tweets plus shallow entity detection and phrase identification. What emerges is words/phrases that have high frequency among the tweets. Thus, all tweets from some location are part of this bag of word and trends are calculated from them. These are called Geographical Trends (GT).

This is essentially a time series, where each instance is of the form: {timestamp, location, [list of trends]}. Thus, each GT-TTL instance includes a list of 10 trends and resembles the top topics of discussion based on tweets coming from the specific geo-locations. We have this data from Nov. 2011 to July 2012, i.e. approximately 8 months. The maximum granularity for GT is 5 minutes.

B. *Brand Audience Trends (BT)*

We also algorithmically track popular accounts on Twitter (e.g., @ESPN, @NYTimes etc.) – these are called 'brands' in the paper. For each of these brand accounts, we collect the tweets of all the users who follow them. So for example, in the case of the New York Times (@NYTimes) we are tracking what their 15.1 million followers are tweeting about. Then we follow a similar algorithm as Twitter to calculate trends from all these tweets. We perform tf-idf and detect entities or phrases that have high frequency. These are trends from the followers/audience of brand New York Times, and thus are called Brand Trends (BT).

In this manner, we collect tweets of followers for approximately 50 Twitter accounts (called brands hereafter). We maintain a diverse category list of brands, including *News* (e.g., @NYTimes), *Sports* (e.g., @ESPN), *Politics* e.g., (@CNNPolitics), *Gaming* (e.g., @IGN), *Entertainment* (e.g., @Miramax) etc. We follow a method similar to that of Twitter when detecting audience trends; i.e. 'trend' is primarily determined by the frequency of word occurrence in the collection of tweets. Thus, after pre-processing, we sort the most frequent words occurring in follower tweets for some brand (Fig. 3). We update this every 2 minutes and therefore, our maximum granularity for BT is 2 minutes. Thus, the BT data set instances also resemble a time series signal in the brand audience world. Each instance is of the form: {timestamp, brand, [list of trends]}.

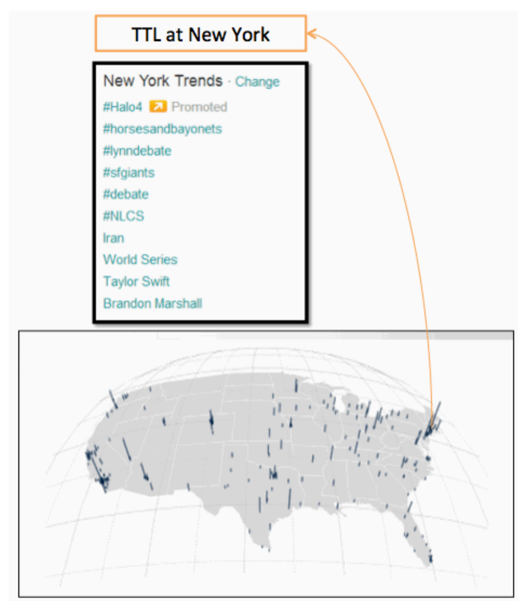

Fig. 2. Example Trending Topic List (TTL) in New York on October 23rd at 8:15 AM - the morning after the 3rd presidential debate in 2012.

C. *Insights from Trend Data*

We assume that the attention of a user group is characterized by the trends that appear within the group. In other words, the attention of user group NY is judged by the trends appearing in TTL of NY. Similarly, trends from audiences following @EA (Electronic Arts) brand indicate behavior of user group (i.e. followers) of EA. With this consideration, we uncovered three key insights from our datasets regarding collective attention in user groups. They are described as follows:

1) User groups possess inherent attention shift tendencies

Different user groups have diverse (unequal) durations for which they can maintain attention on a particular topic. Quick attention shifts is reflected by rapid changes to the TTL of the user group over a contiguous time slots. We noticed that the TTL in some cities (St. Louis) remains constant over multiple hours, whereas in some cities (New York) it changes every 5 minutes.

2) User groups possess intrinsic affinity to certain categories of trends

We also found that user groups in different cities and for different followers are disparately receptive to trends in various categories. For example, San Francisco has strong affinity to trends in *Gaming*, whereas Boston has strong affinity to trends in *Politics*. In a similar fashion, audiences of Pepsi are very attentive to *Entertainment* trends (especially to Justin Bieber), whereas audiences of Burberry have near-zero affinity towards *Sports*.

3) User groups react to real-world events based on a combination of their attention shift patterns and their intrinsic categorical affinity

We chalked some of the major events over a period of 7 months synchronized with the data sets. We found that a real world event first captures the attention of the user group when it aligns to the categorical affinity of the user group and the attention shift tendencies. In other words and contrary to popular belief, it is not more difficult to force something to trend in bigger user groups such as New York compared to Tallahassee, Florida (comparatively smaller user group); provided we know what New York user group has affinity towards and its attention shift tendencies i.e. given the right conditions, trends can break into New York TTL as easily as it does in smaller user groups.

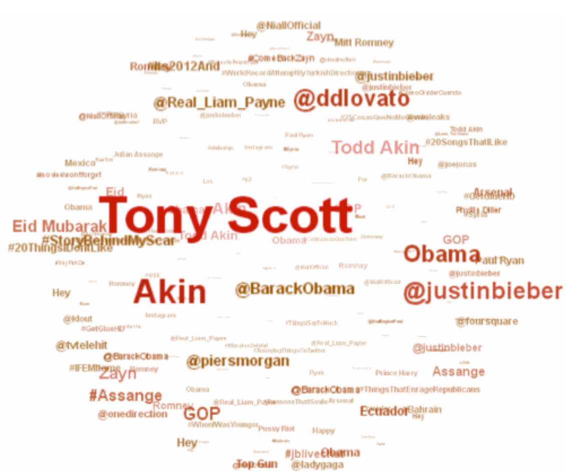

Fig. 3. NYTimes Audience Trends on Aug. 15th, 2012. The bigger/bolder words have higher frequency of occurrence.

One natural question is the amount of overlap observed between the trends appearing through geographical channels (sampled from all tweets at some location) vs. brand channels (calculated from all tweets of some brand's audience). Our experience is that the overlap is less than 7%. This is because the only way audiences of a brand are discussing the same topic/trend as people in that geographical location is when local brands (e.g., the Texas Public Radio) cover some local story (e.g., Texas soldiers death not because of Ebola). The overlap will be higher in smaller cities and for brands that are more local.

IV. The Attention Automaton

We first provide an overview of the automaton in terms of the Twitter ecosystem. We can then explain specifics of measuring attention shift (Section IV. B), calculating categorical affinity (Section IV. C) and formulating the transition probability for the automaton (Section IV. D).

A. Overview

A probabilistic automaton consists of a set of states, a series of actions and a transition probability attached to each potential jump from one state to another based on the action. Let $Distr(X)$ denote the set of all probability distributions over *X*. Then the Attention Automaton (*A*) consists of four components:

1. A set S_A of states.
2. A non-empty set S_A^0 of start states.
3. An action signature $sig_A = (E_A, I_A)$ consisting of external and internal actions respectively. We assume that E_A and I_A be mutually disjoint and the complete set of possible actions is $Act_A = E_A \cup I_A$.
4. A transition relation $\Delta_A \subseteq S_A \times Act_A \times Distr(X)$

14:05:05	15:05:04	16:05:04	17:05:04	18:05:03
#WeStoppedTalkingBecause	#hardesthit	#WeStoppedTalkingBecause	#hardesthit	#OccupyFS
Derren Brown	JLS ARE THE BEST	#iDONTSUPPORT	#icecreamfilms	#masterchef
#insideSBTV	Kate Winslet	#wolves	Villa Park	MCR ARE HEROES
#wolves	#icecreamfilms	Villa Park	#insideSBTV	Finsbury Square
Louise Mensch	MICHAEL JACKSON IS THE KING	MICHAEL JACKSON IS THE KING	Wolves	#HappyTwitterBirthdayLou
Kate Winslet	RIHANNA IS OUR ONLY GIRL	JLS ARE THE BEST	St Paul's	Villa Park
We Need To Talk About Kevin	Tim Minchin	GLEE CAST IS PERFECTION	Swansea	STOP RUINING OUR LIVES
Tim Minchin	Villa Park	RIHANNA IS OUR ONLY GIRL	Paranormal Activity 3	Mario Balotelli
St Paul's	#WBA	Graham Norton	TT's	We Need To Talk About Kevin
The Real Her	St Paul's	Kate Winslet	Anfield	St Paul's

$\gamma_t = Trending\ topic\ list\ at\ time\ instant\ t$

Fig. 4. GT-TTL in London on Oct. 22, 2011. Only hourly TTLs are shown here.

Notice that for each user group, we possess a time series of TTLs. Fig. 4 shows one such time series TTL data for London on Oct. 22, 2011 between 2 PM and 6 PM. Each time-stamped TTL is a state in the automaton. As time passes, the automaton moves to another time-stamped TTL state, based on the new trends that replace some old trends in the latter TTL. We represent the TTL state at time *t* by γ_t. It is also important to note that time is deterministic. At a given instant, the automaton can be in only one state. The start state for all actions on or after time *(t+1)* is γ_t.

A jump from one state to another defines a transition. Each transition is brought upon by an action. The action is a set of impending trends that are attempting to break into the TTL list at time $(t+1)$ so that they can be part of γ_{t+1} (see Fig. 5). Thus, when trend r successfully breaks into the TTL, it fundamentally changes the content of the γ_{t+1} compared to γ_t. The automaton jumping among states based on the action and a transition probability represents changes in the TTL.

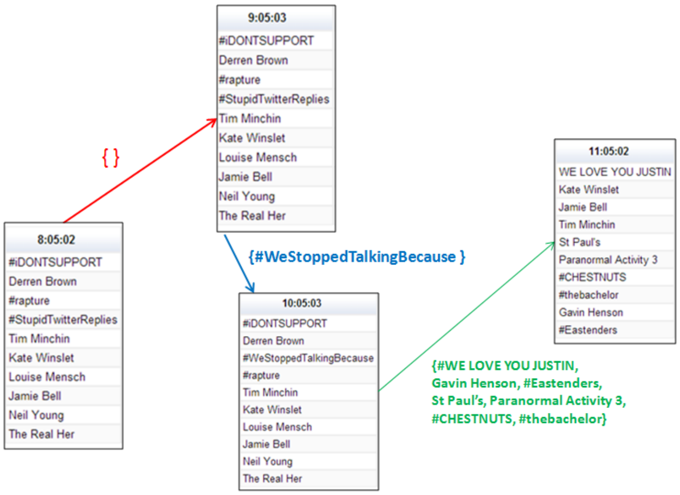

Fig. 5. TTL changes in London on Oct. 22, 2011 between 8-11 AM. Only hourly TTLs are shown here.

The final component of the automaton is the probability of the transition between two states. As mentioned earlier, we found that this probability depends on two factors: (1) the attention shift tendency of the user group, and (2) the categorical affinity of the user group. Section IV. B and IV. C discusses our approach in modeling both these phenomena. Following that, we discuss combining the two attention factors to produce the transition probability in Section IV. D).

Note here that input observations in our problem scenario are in fact, correlated over time. This is why a finite state automaton is a better suited than a Hidden Markov Model (HMM) for modeling the problem. It is well known that HMM disregards the possibility of temporal correlation across inputs features except the defined HMM topology [31].

B. Modeling Attention Shifts

We devise a metric called 'Volatility' to measure the tendency of attention shift over time for a user group. Since we represent content that is receiving attention in a user group based on the TTL in one time slot, measuring the difference between the TTLs in consecutive time slots is an acceptable measure of attention shift.

1) Difference between consecutive TTLs

The difference between consecutive time slot TTLs is basically the edit distance between the two TTLs. In other words, consider each TTL to be a string of trends. Then the difference of two TTLs can be visualized as string-edit distance. We use the Levenshtein distance to measure the difference between two TTLs. Mathematically, the Levenshtein distance between two strings a and b, of sizes i and j, can be expressed as:

$$L_{a,b}(i,j) = \begin{cases} 0 & , i = j = 0 \\ i & , j = 0 \text{ and } i > 0 \\ j & , i = 0 \text{ and } j > 0 \\ \min \begin{cases} L_{a,b}(i-1,j)+1 \\ L_{a,b}(i,j-1)+1 \\ L_{a,b}(i-1,j-1)+[a_i \neq b_j] \end{cases} & , else \end{cases}$$

(1)

As Equation 1 illustrates, $L_{a,b}(i,j)$ is the minimum number of edits required to convert string a to string b. In our scenario, we represent the Levenshtein distance defined as the minimum number of changes needed to convert TTL γ_t to $TTL\ \gamma_{t+1}$ as $L_d(\gamma_{t+1}, \gamma_t)$. By using Equation 1, the Levenshtein distance between TTLs at time 09:05:03 and 10:05:03 in London on Oct. 22, 2011 (shown in Fig. 4) is 2. On the same day, the Levenshtein distance between TTLs at time 17:05:04 and 18:05:03 in London (Fig. 4) is 10.

It is important to note here that we are attempting to measure how much the TTL changes in two consecutive timestamps. We are not concerned if 'similar' trends are part of the TTLs; we only observe if they are exactly the 'same' trends. Thus, edit distances (like Levenshtein distance) is a better choice here than semantic distance (such as WordNet based distances or Google Normalized distance etc.)

The reader might wonder here if it is fair for two TTL lists *A-C-B* and *A-C-D* to have equal distance to the TTL list *A-B-C*? It is true that semantically *A-C-B* and *A-C-D* appear to have different configurations with respect to *A-B-C* and thus their semantic distances from *A-B-C* should be different. However, notice that the automaton is a statistical system. It has limited conceptual knowledge of the semantic difference between *B* and *D*. To it, the states differ by the number of slots that are different, which computationally is the minimum number of single-character edits or the Levenshtein distance. The Attention Automaton, at its current point of development, is a purely probabilistic system. In the future, we may augment the metric with semantic distance.

2) Volatility

Each pair of TTLs in consecutive time slots generates one L_d value. Thus, over a given range of time slots, we have a series of L_d values, which is representative of how fast the TTL of the user group was changing over time. Let us consider T as the number of time slots for the duration of observation, i.e. if we want to calculate the volatility per day and each time slot is 5 minutes, then T= (24*60)/5 = 288.

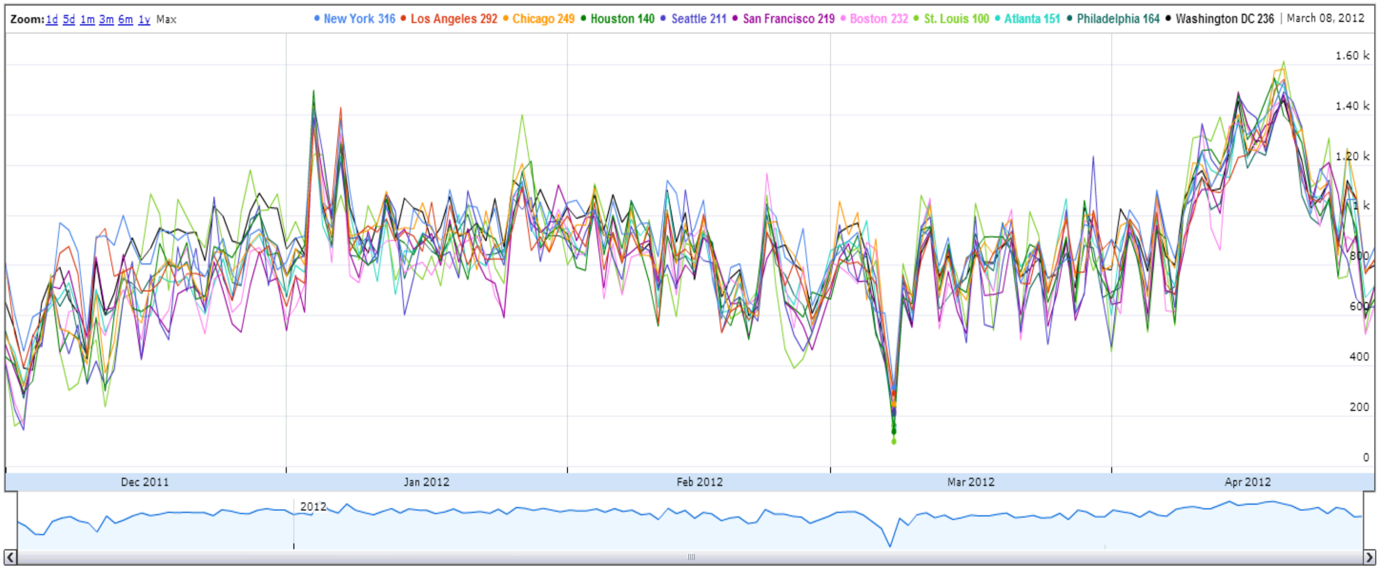

Fig. 6. The volatility per day in some major cities shown here over a period of 5 months. Note the combined minima on March 8th, 2012 across all the cities, which is attributed to the #KONY2012 campaign. The trend captured attention of all cities, and the TTL was almost unchanged on that day.

Then, the volatility for user group *g* starting at time '*st*' can be defined as:

$$Volatility_{st,T} = \sum_{t=1}^{T-1} L_d(\gamma_{t+1}, \gamma_t) \qquad (2)$$

Note that the granularity of volatility measures can be adjusted. For example, if we measure the half-hour-wise change in the TTL for a day, then we can set T to (24*60)/30=48. In Fig. 6, we show the volatility time series over 5 months in some major US cities (with *T*=288). Peaks in the figure refer to days when the TTL was changing rapidly. Peaks are not appealing to us, since rapid changes in TTL indicate that the attention of user groups is shifting rapidly. i.e., there is lack of persistent attention.

Notice that the Volatility metric is not normalized by the number of samples used to calculate it. Therefore, the more the sampling frequency, the higher will be the volatility. In terms of the Volatility Equation 2, $Volatility_{x,T1} \geq Volatility_{x,T2}\ if\ T1 < T2$. This is a good thing, because Twitter is a significantly dynamic system. Two TTLs in an hour gap might appear same, but within the intermediate time slots there might have significant changes in the TTL (trends appearing then disappearing). Said alternately, the lower the sampling frequency, the more smoother the volatility signal will get.

However, a minima in the volatility curve is of significant interest. Minima resemble days when the TTL was not changing significantly. In other words, attention is not shifting constantly instead it is persisting. This can be caused due to two reasons, (1) Nothing is happening that is attention worthy, or (2) Something huge has captured user attention. Especially, when a majority of the user groups display the same minima together on some day, like on March 8th, 2012 (see Fig. 6), then it indicates focused attention within all the user groups to some potentially big event. On March 8th 2012, every city in the US (and most parts of the world) was trending #KONY2012, which was one of the largest online campaigns ever launched through social media[1]. The attention received by the event leads to the combined drop in volatility across all cities on March 8th (Fig. 6).

3) Volatility Signal to Noise (SNR)

From the volatility time series of a user group, we can also infer that some user groups are always volatile (New York), while others are volatile only on few days (Salt Lake City). Volatility Signal-to-Noise ratio (VSNR) is a metric that captures how often an user group is volatile.

Let ϑ_g represent the volatility time series of some user group *g*, i.e.,

$$\vartheta_g = \{(t1{:}\,Volatility_{t1,T}), (t2{:}\,Volatility_{t2,T}), \ldots\}$$

where *t1* is a time instant and *T* is the number of slots over which the volatility was calculated (for a day *T=288*). Then, VSNR can be defined as:

$$VSNR = \delta_g = \frac{Mean(\vartheta_g)}{Std.Dev.(\vartheta_g)} \qquad (3)$$

It is evident why we call this signal-to-noise ratio, since it is basically the ratio of the mean to the standard deviation of the volatility signal [29]. A ratio of mean to standard deviation is the reciprocal of co-efficient of variation, thus serving as an alternative definition of signal to noise ratio. VSNR gives us a single number representing the attention shift tendency of the user group. Fig. 7 depicts VSNR across cities worldwide. We notice Tokyo, New York, Djakarta, London, Los Angeles have high VSNR. In comparison, Montreal, Glasgow, Johannesburg and Mumbai have low VSNR. There can be two explanations of this observation, (1) cities with high VSNR have greater diversity in tweeter profiles - thus lots of topics

[1] Twitter trending #KONY2012 all day in every city significantly contributed to the campaign video receiving a record 60 million views in just 4 days!

capture attention and/or (2) cities with higher VSNR are strongly linked to other user groups, allowing for much larger exposure to diverse information forcing high attention shift.

4) Attention Shift Tendency

Consider the complex ecosystem where a set of user groups (agents) is consuming information. Each user group has some VSNR (δ), indicative of the perturbation of the user group caused by information flow in the underlying social network. Perturbance dynamics in complex networks suggest that there exists a feedback pattern created by the sub-structural network, such that the perturbation of each agent is directly or indirectly affected by another [28].

To put it simply, since the underlying social network governs information flow, the potential of information consumption (attention) of some user group depends on its connections to other user groups through which information reaches it[2]. Assuming all information produced is consumed within the social network, the attention shift tendency of an user group is the probability of consuming information (using attention) by the user group relative to the entire system. It can be defined as a simple ratio:

$$P(\vartheta_{g'}) = \frac{\delta_g{}'}{\sum_{g=1}^{G} \delta_g} \quad (4)$$

where g' is some user group and G is the total number of user groups. Equation 4 gives us the probability of attention shift for some user group g' existing in a world of G groups. A higher probability indicates the user group is potentially more likely to transition to a new state every time.

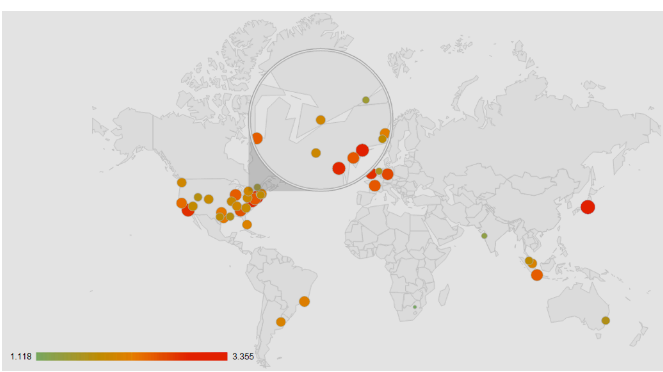

Fig. 7. The Volatility Signal-to-Noise ratio (VSNR) in cities worldwide. Larger circles indicate higher VSNR, representing user groups that are more volatile, while lower VSNR indicates user groups are volatile infrequently.

C. Modeling Attention Shifts

User groups also behave differently to trends in different categories. For example, the audience of Pepsi is highly attentive to any trend about *Entertainment*, especially '`Justin Bieber`', whereas user group of San Francisco is more attentive to trends in *Gaming*, such as the trend '`#halo4`'. We categorize trends over 14 categories, based on whatthetrend.com and the category of the trend word in Wikipedia [25]. These categories are $\mathbb{C} =$ {*entertainment, gaming, lifestyle, science, sports, technology, business, spam, meme, conference or event, news, place or location, holiday or date* and *charity or cause*}.

1) Geographical Trend Initiation

For GT-TTL, a simplistic way to investigate categorical affinity of user groups is to find the location where a trend originated (first trended in the Twitter world) and note the category of that trend. We call this the trend initiation of a user group. Given trends in a category, we can observe the proportion of these trends that originated in some city, and normalize it by initiation in other cities worldwide. This provides us with a Normalized Initiation Score (NIS) between 0 and 1. In Fig. 8, we show the NIS for five major US cities. Notice most *Gaming* trends in the Twitter world originate in San Francisco whereas significant portions of Business trends originate in New York. Somewhat surprisingly, Boston leads all these cities in generating political trends.

2) Follower Affinity

A similar selective affinity to trends is demonstrated in BT-TTL. Followers of specific accounts have selective congeniality to certain trend categories. Fig. 9 shows the categorical distribution of trends that was observed in 3-months worth BT data for followers of four brands, namely Harvard, Burberry, Pepsi and Economist.

It is very interesting to notice how followers are receptive to certain category trends (larger bubbles in Fig. 8) and not so much to others. For example, Pepsi's followers are predominantly sensitive to trends in *Entertainment* whereas Burberry's followers do not care much about *Sports*. Moreover, notice that Harvard followers have a versatile set of categories they are interested in (many same sized bubbles in Fig. 9). Due to lack of space, we cannot provide all the charts. The main indication from this data is that user groups have selective categories they are attentive towards. Therefore, *whether a user group allows an impending trend to enter its TTL is partially dependent on the category of the impending trend.*

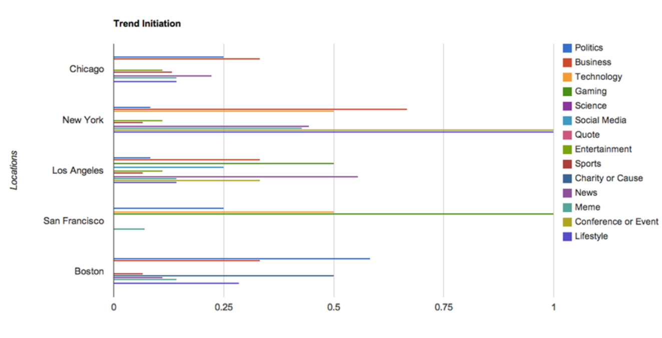

Fig. 8. The normalized trend initiation score worldwide demonstrates the potential of a city to initiate a trend belonging to a particular category compared to other locations worldwide. Data shown here for 5 major US cities.

[2] Observe from Fig. 7 that locations such as NYC, Los Angeles and London have high VSNR, potentially needing to consume the increased information flow attributed to the numerous social network links between users in these locations.

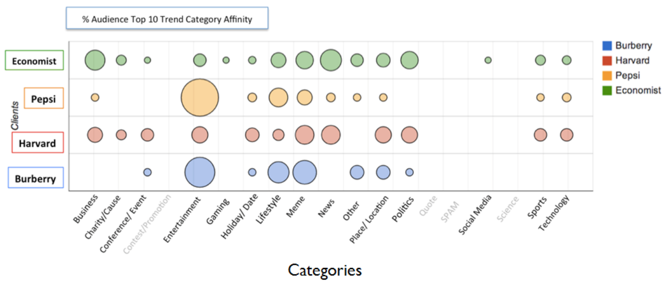

Fig. 9. The distribution of trend categories for followers of brands (Economist, Pepsi, Harvard & Burberry). Size of bubble shows percentage of trends in the user group that belong to a particular category.

3) *User group categorical affinity*

Let the category of trend r be represented as *c(r)*. From the training data, we can also gather the number of trends that appeared in the TTL for the user group g' in some category *c*. Let us denote this by $n_{g'}^{c}$, which is a subset of the total number of trends $N_{g'}$ that appeared in the user group g' over the time period. Consider r' as an impending trend, trying to break into the TTL. Then, the probabilistic categorical affinity of g' to a list of L impending trends can be written as:

$$P(C_{g'}) = \frac{1}{L}\prod_{i=1}^{L}\frac{n_{g'}^{c(r'_i)}}{N_{g'}} \qquad (5)$$

where r'_i is the i-th impending trend, $1 \leq i \leq L$ and $c(r'_i) \in \mathbb{C}$.

D. *Modeling Attention Shifts*

As mentioned earlier, the transition probability determines the probability of a state transition in response to an action stimulus. In the previous sub-sections we describe two probabilistic random variables: attention shift $P(\vartheta_{g'})$ and categorical affinity $P(C_{g'})$ for the user group g'. Note that although $P(\vartheta_{g'})$ has no concern for the action stimulus, $P(C_{g'})$ is fundamentally determined by the action. We know that the joint probability of two mutually independent random variables X and Y is given by:

$$P(X,Y) = P(Y).P(X)$$

In our scenario, the categorical affinity is assumed to be independent of the attention shift tendency, and thus, the transition probability can be considered a joint distribution, written as:

$$Distr(S_A) = P(C_{g'}).\ P(\vartheta_{g'}) \qquad (6)$$

which completes the transition relation described in Section IV. A.

The independence between categorical affinity and attention shift tendency is a valid assumption. This is because the categorical affinity is invariable, whereas the attention shift tendency is dynamic. The attention shift tendency is a variable depending on the information flowing into the user community (external), which is independent of the inherent (internal) affinity of the user community.

Note here that the attention shift tendency of a group depends on other groups, but not necessarily those that have direct information flow among themselves (which is the typical attribute of a 'degree'). Instead, it depends on all the groups in direct and indirect ways (all possible cases). According to perturbation dynamics, the information in the system is constant and will satisfy the countable additive property [40]. In that sense, it is warranted to call it a probability measure.

V. Model Evaluation

We first discuss the evaluation settings, followed by the experimental results. We finally discuss the validity of the model in the real world.

A. *Evaluation Settings*

The experimental settings include preparing the action string to be fed into the automaton, the task to be accomplished, benchmarks and metrics used for evaluation.

1) *Action String Preparation*

To test the model, we first need to prepare a dataset of action trends. This is shortlisted from the trends data. Let all user groups be denoted by U. Let $\overline{\gamma_{g',(t)}}$ represent the trends in the TTL of user group g' at time t. To test a particular user group g', we need to choose actions strings consisting of trends not currently in the user group. Note that these trends are competing simultaneously to be part of $\overline{\gamma_{g',(t+1)}}$. For this purpose, we collect all the unique trends across all user groups U at time t that are not in $\overline{\gamma_{g',(t)}}$. We also record the number of times they have occurred in other TTLs. Thus, this gives us a set D of potential action trends:

$$D = \{d_1, d_2, \ldots, d_m\}$$

where $d_j = (x_j,\ y_j)$ represents a trend along with the number of other TTLs it occurs in at t, i.e. $x \in \{\overline{\gamma_{U,t}} - \overline{\gamma_{g',t}}\}$, $y \geq 1, 1 \leq j \leq m$. We collect the top-$k$ trends in D and choose action strings of different sizes to feed to the automaton. The various action strings encompass the set of actions (mentioned in Section IV.A), which can be written as,

$$Act_A = k_{C_{m*}}, where\ 1 \leq\ m^* \leq 10, K = top_k(D)$$

For our experiments, we choose k =500 and $1 \leq m^* \leq 10$. Note here that each TTL is composed of trends, which is essentially a set. Two consecutive TTLs (with successive time stamps) maybe the same set, or a different set. Let us assume these are sets A and B. The transition from A to B is caused by a set of impending trends M, such that $B - A\ \in M$. In other words, some trends from the set M could replace some trends of the set A, creating the new state B. Now, when we model attention shifts, we have to take into account the temporal nature for the shifts (i.e. the successive timestamps) because the next state (or the next TTL) is not arbitrary in time. In fact, it is the very next the very next time-state of this dynamic system. Thus, the two TTLs are now part of a sequence chained in time.

2) Task Description

The overall purpose of the automaton is to predict most probable future states. The future state depends on the new trends introduced in the next TTL state $\gamma_{g',(t+1)}$. At each time instant t, $Act_{A,t}$ defines the trends that are competing to make it to $\overline{\gamma_{g',(t+1)}}$. However, only q new trends will eventually be in $\overline{\gamma_{g',(t+1)}}$. In other words, $\overline{\gamma_{g',(t+1)}} - \overline{\gamma_{g',(t)}} = q$. The task of our evaluation is to correctly detect the q trends that will cause the automaton to jump from state $\gamma_{g',(t)}$ to state $\gamma_{g',(t+1)}$ forced by the action string of q trends. Said alternatively, we need to detect the q trends, which the automaton will accept out of all the competing trends; that essentially mirrors the actual TTL shift in the Twitter world at that time instant.

3) Benchmarks

Lack of exact comparative work limits our options in selecting benchmarks. However, since this is a time series prediction scenario, we use the traditional *Auto-Regressive Integrated Moving Average (ARIMA)* model which is widely used in statistical analysis of time series with drift [27]. Given a time series, ARIMA can predict future values in the series. The model is generally referred to as an *ARIMA*(a,i,v) model where a, i, and v are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. We use *ARIMA(1,2,1)* to predict trends for future TTLs. The 'statmodels' python package was employed to implement ARIMA in our scenario [26]. Additionally, we use a *random selection scheme*, where the predictor randomly chooses trends to appear in the next TTL. This benchmark is chosen to study if the trend shifts resemble random jumps.

4) Metrics

Since the task is detecting a set of correct trends that mirrors the actual Twitter world TTL state transition, we can simply utilize the precision and recall metrics that are popular in information retrieval. Precision measures how many of the identified trends were actually in q. Recall measures how many of the q trends were retrieved. The harmonic mean of precision and recall is called *F-Score*, which is $\frac{2.precision.recall}{(precision+recall)}$ and serves as our evaluation metric. A higher F-Score suggests better performance.

B. Results

The training phase involves using a prior section of the time series in order to gain information about the user group's attention shift tendency and categorical affinity. Both Equation 4 and Equation 5 are supported by this training data. Therefore, we can look at a few months' data for training and predict the future months in testing. We perform separate experiments on the two datasets, GT-TTL and BT-TTL time series. Experiments undergo a 10 fold cross-validation, results of which are discussed below.

1) User groups by geographical locations (GT)

We randomly select 30 locations worldwide to perform these tests. The average F-score obtained using the Attention Automaton, ARIMA and Random models was 0.49, 0.34 and 0.18 respectively. Overall, the *F-score performance using the Attention Automaton was 44% better than ARIMA and 171% better than random selection*. The F-score of user groups in 10 out of the 30 locations chosen for testing is provided in Table I, which was generated using 3 months of the user group data for training and 3 months for testing. Corresponding Precision and Recall values are also provided in Table I and II.

Since our data is based on a 6-month period, we could take different combinations of test and training data size to investigate the performance of the automaton compared to the benchmarks with varying exposure to data, i.e. varying the data available for training phases, which will essentially vary the knowledge the automaton has of attention shifts and categorical affinity. Results of this experiment is reported in Fig. 10, where the size of the bubble indicates the value of the F-Score and the test and training values refer to months of the time series used. The models perform differently based on the size of the training data they have been exposed to. For example, the random technique does not depend on training data, but the more data we test the random model on, the poorer its performance becomes (the bubbles keep growing smaller when the test data size is increased). The Attention Automaton has better performance when it can see more of training data, since from Fig. 10 we can observe that the bubble sizes are bigger every time it sees 3 months of training data. The ARIMA usually performs better when it is exposed to more training data than test data. In all cases, the Attention Automaton outperforms the other two benchmarks.

2) User groups by brand following (BT)

We select 30 user groups of brands to perform similar experiments as the GT user groups. For each BT user group, half the time series is used for training and the other half for testing. Results are reported in Table II for ten of these brand user groups. The average F-Score achieved using the random method, the ARIMA model and the Attention Automaton is 0.163, 0.317 and 0.552 respectively. *Overall, the improvement using the Attention Automaton with respect to F-score was 238% over the random scheme and 74% over the ARIMA model*.

TABLE I

F-scores obtained in testing different models on user groups from some major **cities/geolocations** in the world on Twitter. The results are the averages of 10 random repeats along with their standard deviations on F-score. All three methods are tuned with 10-fold cross-validation.

User Group	Random				ARIMA				Automaton (proposed)			
	FScore	Precision	Recall	STD	FScore	Precision	Recall	STD	FScore	Precision	Recall	STD
New York	0.19	0.15	0.25	±0.02	0.29	0.25	0.34	±0.01	***0.42***	*0.41*	*0.43*	±0.02
Los Angeles	0.17	0.11	0.34	±0.01	0.30	0.21	0.5	±0.01	***0.46***	*0.56*	*0.39*	±0.01
Baton Rouge	0.14	0.11	0.21	±0.01	0.38	0.33	0.44	±0.01	***0.53***	*0.54*	*0.52*	±0.03
Boston	0.19	0.12	0.44	±0.01	0.34	0.38	0.31	±0.02	***0.44***	*0.37*	*0.55*	±0.01
Paris	0.15	0.18	0.13	±0.02	0.36	0.50	0.28	±0.02	***0.40***	*0.64*	*0.29*	±0.02
London	0.14	0.10	0.23	±0.01	0.27	0.28	0.26	±0.01	***0.43***	*0.47*	*0.4*	±0.01
Dublin	0.16	0.10	0.35	±0.01	0.37	0.33	0.45	±0.03	***0.56***	*0.80*	*0.43*	±0.02
Atlanta	0.18	0.13	0.29	±0.01	0.35	0.28	0.45	±0.01	***0.55***	*0.71*	*0.45*	±0.03
San Francisco	0.13	0.09	0.21	±0.01	0.33	0.29	0.39	±0.02	***0.40***	*0.34*	*0.48*	±0.02
Glasgow	0.19	0.13	0.38	±0.02	0.35	0.25	0.34	±0.02	***0.48***	*0.68*	*0.37*	±0.02

One interesting observation is that the F-score improvement of the Automaton over ARIMA is different for different user groups. More precisely, Automaton performs 61% better for user groups of EA, Pepsi, Burberry and Wal-Mart compared to Harvard, Associated Press or CNN. We contribute this nature to the distribution of categorical affinity of user groups. User groups of Pepsi, EA and Burberry have small number of categories they have affinity towards, effectively reducing the decision space for prediction. As shown in Table II, Pepsi has high affinity to '*Entertainment'*. In contrast, user groups of CNN/Harvard have a large number of categories they have affinity towards. ARIMA lacks understanding of categorical affinity, as it is driven by the statistical variation in the time series. Therefore, for user groups that have very high affinity to very few categories, the Attention Automaton performs significantly better than ARIMA.

C. *Validity of the Attention Automaton in the Real World*

In this section, we attempt to validate our modeling criterion of volatility and categorical affinity with respect to how user groups react to real world events. The following real-world events in 2012 were chosen for this experiment (category of event is shown in parenthesis):

1. April 8th: Romney wins three primaries against Santorum. (*Politics*)
2. April 11th: Earthquake (*News*)
3. April 14th: Syria Protests (*News*)
4. April 19th: Bahrain Protests (*News*)
5. March 6th: Hollande defeats Sarkozy in French elections. (*Politics*)
6. May 11th: Justin Bieber song 'Turn to you' releases. (*Entertainment*)
7. May 23rd: Justin Bieber 'Believe' tour dates released. (*Entertainment*)
8. June 17th: Rodney King dies. (*Cause*)
9. June 19th: Justin Bieber releases new album 'Believe'. (*Entertainment*)

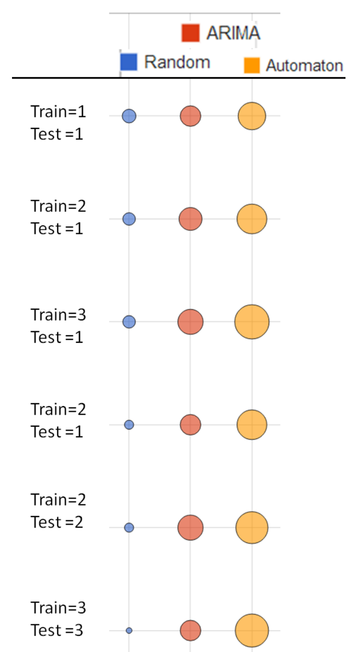

Fig. 10. Variation of model performance with varying duration of data considered for the training and testing phases. Train = 3 means trained on three months of data. Similarly, Test = 1 implies tested on 1 month of data.

As explained earlier, a drop in volatility (minima) hints at focused attention of the user group. We take note of when the volatility drops in a user group and what corresponding event occurred on that day. We then check the trends in the user group on that day and draw a correlation to whether the user group trends reflected the real world event.

Within user groups of brand audiences, we noticed a strong relation between our attention model and the user groups' reaction to these physical world events. This phenomenon is described in Fig. 11, where we plot the volatility of a brand user group (Y-axis) for three months of trends data (X-axis). We observed the following user groups during this experiment: *Financial Times, Harvard, Harvard Med, Walmart, Pepsi, American Museum of Natural History (AMNH), NY Times, Associated Press, eHealth* and *Economist*.

Fig. 11 shows the variation of volatility in each user group as events 1-9 (described above) transpired in the physical world. Significant drops in volatility indicate focused attention (user group which has significant volatility drop for some event is marked beside the minima in Fig. 11). Notice how Pepsi user group pays attention to anything involving Justin Bieber (*Entertainment*). The categorical affinity within user groups, which we observed in Fig. 9, is triggered in this real world situation, causing audiences of Pepsi to strongly react to *Entertainment* events. Similarly, user group of Associated Press is very attentive to *Politics* and *News*. Audience of NY Times is attentive to trends in *Cause* category, with respect to the death of Rodney King. All user groups pay focused attention when there is an earthquake (emergency breaking news)! Therefore, a volatility based attention shift and categorical affinity proves valuable in modeling the reaction of user groups in response to physical world events. This validates the soundness of the Attention Automaton in modeling the reaction of Twitter communities in response to physical world events.

VI. Conclusion and Future Work

Social multimedia is dwarfing traditional content and media and setting new boundaries to web applications and disruptions in business strategies around it. We know that an exorbitant amount of traffic on the web is actually social multimedia traffic [33]. This is causing our lives to be tied to social multimedia. For example, YouTube reaches more US adults below 34 than any cable network. 1 million websites have been integrated into Facebook. 210,000 years of music have been played on Facebook. Even news production uses tweets to add a voice of the people in addition to the experts in the studio. The task of measuring user attention in response to so much social activity is very important for several web applications. Moreover, the computation of such attention is vital so that automated systems can keep measuring it and alert other systems based on where attention focuses.

In this paper, we present a probabilistic automaton for quantifying the attention of social network communities (user groups). Two types of communities are tested here: geographical location based and brand follower based. Our model demonstrates that we can measure two key aspects of user groups: (1) their attention shift tendency, and (2) their categorical affinity for trends in selected categories.

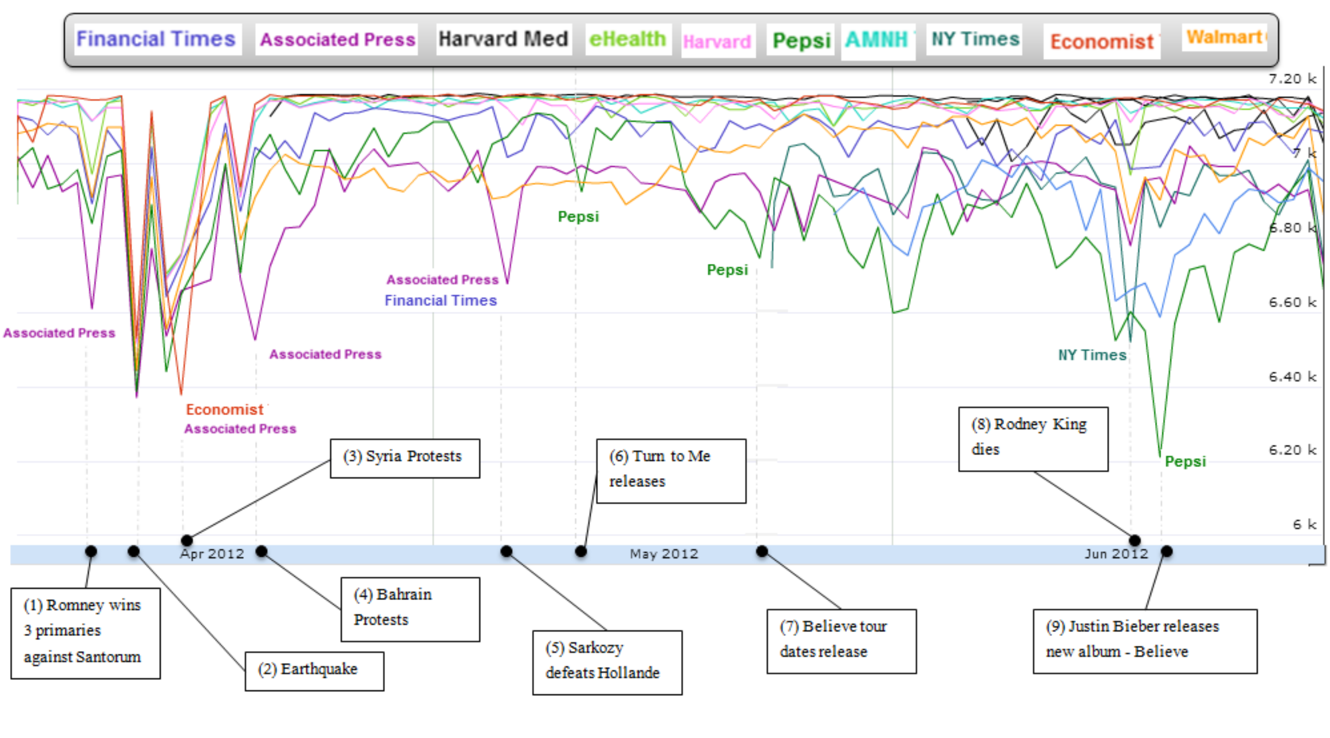

Fig. 11. The plot shows variation of user group attention in response to real world events. Described in Section V.C, this validates that dynamic volatility, a key component in the Attention Automaton, is able to correctly reflect the categorical attention of user groups in brands in real world scenarios.

TABLE II
F-scores obtained in testing different models on user groups of following major **brands/accounts** in the world on Twitter. The results are the averages of 10 random repeats along with their standard deviations for F-scores. All three methods are tuned with 10-fold cross-validation.

User Group	Random				ARIMA				Automaton (proposed)			
	FScore	**Precision**	**Recall**	**STD**	**FScore**	**Precision**	**Recall**	**STD**	**FScore**	**Precision**	**Recall**	**STD**
CNN	0.18	0.15	0.22	±0.01	0.29	0.27	0.32	±0.01	***0.42***	0.54	0.34	±0.01
EA	0.17	0.12	0.3	±0.01	0.30	0.27	0.34	±0.03	***0.46***	0.54	0.4	±0.02
Burberry	0.14	0.10	0.26	±0.01	0.38	0.37	0.39	±0.01	***0.53***	0.69	0.43	±0.02
Economist	0.19	0.20	0.18	±0.01	0.34	0.38	0.31	±0.02	***0.44***	0.52	0.38	±0.01
Financial Times	0.15	0.12	0.21	±0.02	0.36	0.32	0.41	±0.02	***0.40***	0.51	0.33	±0.04
Harvard	0.14	0.10	0.24	±0.01	0.27	0.25	0.3	±0.02	***0.43***	0.45	0.41	±0.01
Pepsi	0.16	0.11	0.31	±0.02	0.37	0.41	0.34	±0.02	***0.56***	0.65	0.49	±0.04
Walmart	0.18	0.13	0.28	±0.02	0.35	0.44	0.29	±0.03	***0.55***	0.73	0.44	±0.03
Associated Press	0.13	0.10	0.19	±0.01	0.33	0.34	0.32	±0.01	***0.40***	0.44	0.37	±0.03
NYTimes	0.19	0.14	0.31	±0.02	0.35	0.31	0.4	±0.01	***0.48***	0.60	0.4	±0.02

The Attention Automaton uses these measures to compute the probability of new/impending trends that will receive attention from some user group. Our work is a step forward towards building a predictive computable model of attention and the proposed approach accomplishes more than 40% improvement over benchmarks. It is well known that the unit of marketing is attention, thus, the Attention Automaton has significant potential in boosting marketing and advertising applications.

There is a lot of scope in this line of research. Social networks are rich in signals about the human condition (e.g., *what we want*), a primary understanding of which is the cornerstone of advertising. We use trends as a feature of attention in our work, but it would be interesting to explore what other properties of social networks could be used for the same. Our work is also limited by the accuracy of categorical classification of trends. This could be especially challenging for trends beyond hash tags; trends such as multi-word expressions like '*TeamKhleoThomasShoutouts*' or '*Stop judging Justin Bieber'* or mixed category trends like *'Brazil loves Justin Bieber*' (categories: *location* + *Entertainment*). Moreover, it would be fascinating to understand how exactly trends compete to break into the TTL in terms of game theory.

Twitter has often been compared to the 'pulse' of the online social network world [23]. It has been leveraged to build remarkable applications, from Olympics sentiment analysis and cross-domain media recommendations to tracking political opinions and flu trends [22, 24]. All such applications utilize the Twitter signal as a pulse for the millions of online users. The Attention Automaton goes one step further - it models what sets their pulse racing.

REFERENCES

[1] B. A. Huberman, D. A. Romero and F. Wu, "Social networks that matter: Twitter under the microscope," in *Computing Research Repository - CORR*, vol. abs/0812.1, no. 1, 2008.

[2] C. Wagner, M. Rowe, M. Strohmaier and H. Alani, "What Catches Your Attention? An Empirical Study of Attention Patterns in Community Forums," In *Inter. Conf. on Weblogs and Social Media* 2012.

[3] L. Backstrom, E. Bakshy, J. Klienberg, T. M. Lento and I. Rosenn, "Center of attention: How facebook users allocate attention across friends," In *Inter. Conf. on Weblogs and Social Media* 2011.

[4] Janette Lehmann, Bruno Gonçalves, José J. Ramasco, and Ciro Cattut., "Dynamical classes of collective attention in twitter," In *Proceedings of the 21st international conference on World Wide Web* (WWW '12).

[5] R. Hanna, A. Rohm and V. L. Crittenden, "We're all connected: The power of social media ecosystem," In *Business Horizons special issue on Social Media*, Vol. 54, Issue 3, pp. 265-273.

[6] Eytan Bakshy, Dean Eckles, Rong Yan, and Itamar Rosenn. "Social influence in social advertising: evidence from field experiments," In *Proceedings of the 13th ACM Conference on Electronic Commerce* (EC '12).

[7] Ayman Farahat and Michael C. Bailey, "How effective is targeted advertising?," In *Proceedings of the 21st international conference on World Wide Web* (WWW '12). ACM, New York, NY, USA, 111-120.

[8] N. O. Hodas and K. Lerman, " How Visibility and Divided Attention Constrain Social Contagion," In *SocialCom 2012.*

[9] Mao Ye, Thomas Sandholm, Chunyan Wang, Christina Aperjis, and Bernardo A. Huberman, "Collective attention and the dynamics of group deals," In *Proceedings of the 21st international conference companion on World Wide Web* (WWW '12 Companion). ACM, New York, NY, USA, 1205-1212.

[10] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. "What is Twitter, a social network or a news media?," In *Proceedings of the 19th international conference on World Wide Web* (WWW '10). ACM, New York, NY, USA, 591-600. 2010.

[11] Li Yujian and Liu Bo, "A Normalized Levenshtein Distance Metric," In *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 6 (June 2007), 1091-1095. 2007.

[12] H. A. Simon, "The Sciences of the Artificial (3rd ed.)," Cambridge, MA: *The MIT Press*, ISBN 0-262-69191-4. 1996.

[13] T. H. Davenport and J. C. Beck, "The Attention Economy: Understanding the New Currency of Business," *Harvard Business School Press*. ISBN 1-57851-441-X. 2002.

[14] S. Xu, Y. Zhu, H. Jiang and F. C. M. Lau, "A user-oriented webpage ranking algorithm based on user attention time," In *Proc. of the 23rd AAAI Conference on Artificial Intelligence*. 2008.

[15] P. P. Zubcsek and M. Sarvary, "Advertising to a social network," In *Quantitative Marketing and Economics,* Vol. 8, No. 1, pp. 71-107. 2011.

[16] Arpita Ghosh and Preston McAfee, "Incentivizing high-quality user-generated content," In *Proceedings of the 20th international conference*

on World Wide Web (WWW '11). ACM, New York, NY, USA, 137-146. 2011.

[17] B. A. Huberman, “Social dynamics in the age of the web,” In *ACM Conf. on Hypertext*, pp. 1-2, 2008.

[18] L. Weng, A. Flammini, A. Vespignani and F. Menczer, “Competition among memes in a world with limited attention,” In *Nature*, Scientific Reports 2, Article No. 335. 2012.

[19] Sara Noeman and Amgad Madkour, “Language independent transliteration mining system using finite state automata framework,” In *Proceedings of the 2010 Named Entities Workshop* (NEWS '10). Association for Computational Linguistics, Stroudsburg, PA, USA, 57-61. 2010.

[20] Azaria Paz, “Introduction to Probabilistic Automata (Computer Science and Applied Mathematics),” *Academic Press, Inc.,* Orlando, FL, USA. 1971.

[21] P. Dupont, F. Denis, and Y. Esposito, “Links between probabilistic automata and hidden Markov models: probability distributions, learning models and induction algorithms,” *Pattern Recognition* 38, 9 (September 2005), 1349-1371. 2005.

[22] S. D. Roy, T. Mei, W. Zeng and S. Li, “SocialTransfer: Cross-Domain Transfer Learning from Social Streams for Media Applications,” *In Proc. of ACM Multimedia (ACM MM)*, Nara, Japan, 2012.

[23] Nicholas A. Diakopoulos and David A. Shamma, “Characterizing debate performance via aggregated twitter sentiment,” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10). ACM, New York, NY, USA, 1195-1198. 2010.

[24] Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner and Isabell M. Welpe, “Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment,” In *Inter. Conf. on Weblogs and Social Media* 2010.

[25] Kathy Lee, Diana Palsetia, Ramanathan Narayanan, Md. Mostofa Ali Patwary, Ankit Agrawal, and Alok Choudhary, “Twitter Trending Topic Classification,” In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops* (ICDMW '11). IEEE Computer Society, Washington, DC, USA, 251-258. 2011.

[26] http://pypi.python.org/pypi/scikits.statsmodels

[27] Mills, Terence C., “Time Series Techniques for Economists,*” Cambridge University Press*. 1990

[28] Marshall A. Kuypers, Walter E. Beyeler, Robert J. Glass, Matthew Antognoli, and Michael D. Mitchell, “The impact of network structure on the perturbation dynamics of a multi-agent economic model,” In *Proceedings of the 5th Inter. Conf. on Social Computing, Behavioral-Cultural Modeling and Prediction* (SBP'12). 2012.

[29] J. T. Bushberg, “The Essential Physics of Medical Imaging,” in *(2e) Philadelphia: Lippincott Williams & Wilkins*, p. 280. 2006.

[30] Gilad Lotan, “Data Reveals that Occupying Twitter Trending Topics is Harder than it Looks,” http://blog.socialflow.com/post/7120244374/data-reveals-that-occupying-twitter-trending-topics-is-harder-than-it-looks. 2010.

[31] Bourlard, Hervé, and Samy Bengio, "Hidden markov models and other finite state automata for sequence processing," In *The Handbook of Brain Theory and Neural Networks*, (2002).

[32] Suman Deb Roy, Tao Mei, Wenjun Zeng, and Shipeng Li, "Empowering Cross-Domain Internet Media with Real-Time Topic Learning from Social Streams," In *Multimedia and Expo (ICME), 2012 IEEE International Conference on*, pp. 49-54. IEEE, 2012.

[33] Suman Deb Roy, Tao Mei, Wenjun Zeng, and Shipeng Li, "Towards cross-domain learning for social video popularity prediction," Multimedia, IEEE Transactions on 15, no. 6 (2013): 1255-1267 .

[34] Suman Deb Roy, Gilad Lotan, and Wenjun Zeng, "Social multimedia signals: Sense, process, and put them to work," *MultiMedia, IEEE* 20, no. 1 (2013): 7-13.

[35] Kumar, S., Morstatter, F., & Liu, H, “Twitter data analytics”. Springer, 2014.

[36] D. Lazar, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, “Computational Social Science,” *Science*, Vol 323, Iss 5915 : 721-723, 2009.

[37] T Preis, D Reith, HE Stanley, “Complex dynamics of our economic life on different scales: insights from search engine query data,” In *Philosophical Transactions of the Royal Society,* 2010.

[38] Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., & Preis, T, “Quantifying Wikipedia usage patterns before stock market moves,” In *Scientific reports*, *3*, 2013.

[39] Maksim Kitsak, Lazaros K Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, Hernán A Makse., “Identification of influential spreaders in complex networks”. In *Nature Physics*. Vol. 6 Issue 11, pp: 888-893, 2010.

[40] Haken, H, “Information and self-organization: A macroscopic approach to complex systems,” In *Springer Science & Business Media*, 2006.