Learning a Linear Influence Model from Transient Opinion Dynamics

Abir De
IIT Kharagpur, India

Sourangshu Bhattacharya
IIT Kharagpur, India

Parantapa Bhattacharya
IIT Kharagpur, India

Niloy Ganguly
IIT Kharagpur, India

Soumen Chakrabarti
IIT Bombay, India

ABSTRACT

Many social networks are characterized by actors (nodes) holding quantitative opinions about movies, songs, sports, people, colleges, politicians, and so on. These opinions are influenced by network neighbors. Many models have been proposed for such opinion dynamics, but they have some limitations. Most consider the strength of edge influence as fixed. Some model a discrete decision or action on part of each actor, and an edge as causing an "infection" (that is often permanent or self-resolving). Others model edge influence as a stochastic matrix to reuse the mathematics of eigensystems. Actors' opinions are usually observed globally and synchronously. Analysis usually skirts transient effects and focuses on steady-state behavior. There is very little direct experimental validation of estimated influence models. Here we initiate an investigation into new models that seek to remove these limitations. Our main goal is to estimate, not assume, edge influence strengths from an observed series of opinion values at nodes. We adopt a linear (but not stochastic) influence model. We make no assumptions about system stability or convergence. Further, actors' opinions may be observed in an asynchronous and incomplete fashion, after missing several time steps when an actor changed its opinion based on neighbors' influence. We present novel algorithms to estimate edge influence strengths while tackling these aggressively realistic assumptions. Experiments with Reddit, Twitter, and three social games we conducted on volunteers establish the promise of our algorithms. Our opinion estimation errors are dramatically smaller than strong baselines like the DeGroot, flocking, voter, and biased voter models. Our experiments also lend qualitative insights into asynchronous opinion updates and aggregation.

1. INTRODUCTION

Opinion formation and its propagation is a crucial phenomenon in any social network and has been studied widely both from a computational perspective as well by sociologists and psychologists. One of the ways opinion is thought to be formed (and hence propagated) is through a process called *informational influence*, where a user forms her opinion about a topic by learning information about that topic from her neighbors. Research on opinion propagation predominantly follows two kinds of models: discrete [5, 1, 16], where opinions are quantized/ordinal, and continuous [7, 10, 11]. A recent paper [6] seeks hybrid models.

CIKM'14, November 3–7, 2014, Shanghai, China.

ACM 978-1-4503-2598-1/14/11 ...$15.00.
http://dx.doi.org/10.1145/2661829.2662064 .

1.1 Limitations of prior work

Despite these advances, current understanding of opinion and influence is limited in several ways. Many models assume that the weights of influence of the neighbors are identical, or set them arbitrarily[5, 10, 16], without regard to observed behavior. Another implicit assumption made by many models [7, 10, 11] is that opinions converge and/or consensus or polarization is reached as a steady state. However, if one carefully looks into the recent scenario, information is continuously fed into the system and opinion of people continuously unfurl. So unlike the previous setting where most of the analysis is based on the steady state behavior of the system, the apparently long *transient* behavior of realistic systems need to be studied.

The feasibility and need for studying the *transient* behavior has arisen from the large amount of user generated content, e.g. tweets, which are now available for analysis. Subsequently, market survey has become a continuous feature, for example, people's rating of leaders are collected almost every month rather than just before election, sentiment of people can be continuously assessed from the comments/tweets they post. In such a setting it is now necessary to learn the underlying opinion dynamics.

1.2 Our contributions

We initiate investigation into the following type of influence model. Agents have a quantitative opinion. They influence each other through edges of the social network. Influence is linear in nature, but not necessarily stochastic. In direct contrast to the majority of prior work, influence strengths along edges are not known a priori. Instead, the state of opinion of different actors are observed at various sampled instants of time. Our goal is to learn the parameters of the underlying linear opinion propagation model. Note that our focus is on estimation errors of individual influence edges, rather than aggregate behavior like bifurcation.

In another departure from prior work, we do not assume that these observations are collected at regular intervals, or at the same instants across all actors. Collecting the opinion of an individual as soon as it is updated (push mode) would capture the most information for influence estimation. However, in a practical setting, it may be only possible to collect it periodically (pull mode). E.g., opinion about political leaders may be collected in monthly surveys. But the number of updates in between may vary widely across actors and time. E.g., people may update their opinions much more fre-

quently before and during an election. We also do not make any assumptions about convergence to consensus, polarization, or fragmentation. The observations may well include ones made during transient stages.

To the best of our knowledge, this is the first attempt at learning linear opinion propagation dynamics from observed opinion values of the individual agents without appealing to steady state behavior.

1.3 Experimental validation

In addition to our theoretical investigation, we report on a series of experiments with five data sets to validate our influence model and estimation algorithms. Three of these were collected by running controlled, in-house, social opinion exchange processes. Here we attempted to capture every opinion change of all participants, who were told to form opinions base solely on discussions with designated social network neighbors. This data will be made available. Two further data sets were collected from Reddit and Twitter. Here we encountered the more general and realistic setting where opinion values are collected periodically, but opinion propagation may take place aperiodically.

Compared to several well-known baselines such as the voter model, the biased voter model, the flocking model, and DeGroot's linear model, the models we propose here yield significantly smaller prediction error, smaller by factor of 2–15. Tuning certain regularization parameters in our model shows clear signs of meaningful edge influence learning. We also present analysis of influence propagation time horizons based on our models.

2. RELATED WORK

While they are often used interchangeably, Merriam-Webster defines an opinion as "a conclusion thought out yet open to dispute" and a sentiment as "a settled opinion reflective of one's feelings" [19]. In this work, we use the term opinion, in view of their dynamic nature.

Not all influence is propagated along social network edges; external events also impact agents. However, Myers *et al.* [18] developed a detailed model for blending external and social influence, and found that 71% of the information transfer volume (suitably characterized) in Twitter can be attributed to information diffusion across social links. Here we will focus exclusively on influence conveyed by social links.

2.1 Discrete opinion

Discrete models assume that the opinions are discrete (binary or ordinal/quantized). The voter model [5] belongs to this category. At each step, a node is selected at random; this node chooses one of its neighbors uniformly at random (including itself) and adopts that opinion as its own. This model always lead to consensus which is rare in many social scenarios. A modified version of the voter model is called label propagation [25] where the node adopts the majority opinion from among its neighbors. These models are limited to achieve consensus.

One way to overcome such limitations is to incorporate stubborn agents [24]. Another way [1] is to have each agent adopt its neighbors' opinion depending on the similarity with her own. This model leads to polarization instead of consensus. This was entirely a data-driven study with no rigorous analysis. A further unifying variation was analyzed by Lanchier [16]. In that model, an agent adopts another agent's opinion if those opinions are within a certain distance or difference called the confidence threshold. Lanchier showed that small (large) threshold values lead to polarization (consensus) with high probability. Kempe *et al.* [13] brought forward the concept of influence-selection whereby an agent is not only influenced by other agents which has similar opinion but also selects for interaction agents who are similar to itself. They proved that such behavior can stabilize over arbitrary graphs and precisely characterize the set of all stable equilibria.

Discrete opinions are a natural model for some applications, but not others. E.g., opinion about world population at a future date, or the concentration of atmospheric CO_2, or the number/fraction of votes a politician might get, are all effectively continuous.

2.2 Continuous opinion

Our present work is in the other category of continuous opinions. Many models for continuous opinion assume, like us, that neighbors influence *linearly* the opinion of an agent [7], reaching limited consensus. Analysis is frequently grounded in the mathematics of matrix eigensystems, physics and theoretical biology. They are based, for example, on bird flocking models [10] and Cellular Automata [11]. In the flocking model, a node i (agent) with opinion x_i first selects the set of neighbors j having opinions x_j so that $|x_i - x_j| \leq \epsilon$, and then updates its own opinion by averaging them.

There is also a large body of work (see [17, 3] and references therein) that has sought to characterize the convergence of bounded confidence dynamics to either absolute consensus or some clustering (polarization). But not all papers focus on convergence. Bindel *et al.* [2] state that in many social settings consensus may never be attained. They characterize the cost of disagreements in a game-theoretic setting.

Of course, there are other occasions where only a discrete opinion model will fit, and network averaging in the continuous sense is not meaningful [4]. Agents must choose from a fixed discrete set of options. Various formulations of graphical games showed that characterizing stability even for a two-strategy game is very difficult.

We chose continuous opinions to enable some theoretical handle on our newly-introduced complications such as possible transience and asynchronous observations. However, there are some important distinctions with earlier work. DeGroot [7] assumed a row-stochastic influence matrix with $w_{ij} \geq 0$, and opinions in the range $[0, 1]$ (which stochastic updates preserved). In our case, opinions can be unbounded, updates are no stochastic (influence can be negative, and an agent's combination rule is not convex), and zero is a special opinion value separating two polarities of opinion.

2.3 Hybrid models

A very recent paper [6] proposes a hybrid model between discrete and continuous. It proposes a *biased voter model*, which is a unification of the voter model with flocking. Each agent is driven by a mix of three forces: stubbornness (ignoring others' opinions), DeGroot's permissive averaging with neighbors, and biased conformance, which chooses influencing agents biased toward those whose opinions are already somewhat close to that of the base agent. A preliminary data study is used to justify the tension between these forces, and the resulting model is analyzed to the following two ends. First, even if an individual agent changes opinion continually, the relative frequencies of different opinions converge. Second, consensus still happens under certain conditions. This paper is not concerned with influence estimation on individual edges, which is our main goal.

2.4 Maximizing influence

Yet other works [21, 14] assume fixed topology and edge weights or propagation rules, and seek to select an initial set of active (or 'infected') so as to maximize some kind of cascading effect to the rest of the network. We do not seek to maximize influence; we *ob-*

serve a dynamic influence process and estimate influence strength of all edges.

2.5 Estimating edge influence strength

The vast majority of the work discussed above assume some kind of fixed influence strength on each edge. A notable exception [8], which, however, returns to the domain of some discrete action on part of one agent, that precipitates the same action in another agent at some subsequent time. Given the temporal ordering, influence propagation is acyclic, an assumption at odds with any kind of reciprocal, continual influence. But this simpler setup allows them to $p_{v,u}$ from a form of soft-OR influence model at each node: $p_u(S) = 1 - \prod_{v \in S}(1 - p_{v,u})$, where S is the set of neighbors of u that have already committed the action, and $p_u(S)$ is compared to a threshold to decide if u should also commit it. Another notable example of influence estimation is by Shahrampour *et al.* [22], who provide a purely theoretical analysis of the online continuous case, but do not deal with asynchronous observations, or validate on real data.

3. MODEL AND METRICS

In this section, we describe a new framework for modeling opinion dynamics on a social network. Our choice of model is driven by study of short-term or transient behaviour of opinions in a social network. In Section 3.1, we propose a general linear model for opinion propagation, and then, in Section 3.2, we propose metrics for data-driven evaluation of opinion models using observed data. In the next Section 4 we will get to the parameter-learning algorithms, the central contribution of this paper.

3.1 Model Definition

DeGroot [7] defines opinion as the "subjective probability" a person assigns to an event. Thus, opinion of each person on a topic is a real number between 0 to 1 in his framework. The DeGroot model proposes opinions of a person after one round of 'discussion' with others, to be a *convex* combination of others' opinions with weights assigned to each person. DeGroot described conditions under which persons discussing and maintaining these constraints will asymptotically reach consensus.

In a departure from DeGroot, we define opinion as an *arbitrary real number* describing a person's opinion / sentiment on an issue, real world event, product, etc. Our notion of opinion is more akin to opinion mining or sentiment analysis (see e.g. Pang et al. [19]), where the polarity (+ve or $-$ve) and magnitude of the opinion are important. For example, on a recently launched product, an opinion value of $+1$, 0 and -1 could mean that the product is "good", "neutral" and "bad" respectively. Thus, for our model 0 (zero) becomes a natural threshold between "good" and "bad". Note that we are not imposing any bounds on the opinion values in the model, which allows us to define dynamics using arbitrary *linear* combination of other opinions, rather than just *convex* combinations.

We denote the opinion of a person i at time instant k as $x_k^i \in \mathbb{R}$. Let $G = (V, E)$ be a directed graph representing a social network where V is the set of vertices or nodes representing people who are forming and propagating opinions. Opinions are propagated only through the set of edges: $E \subseteq V \times V$, which represent connections between people. This is justifiable on many modern social networks as the platform allows posting messages only to the neighbors. Also, let $N = |V|$ be the total number of people in the social network.

In this paper, we propose a *linear opinion propagation model*: opinion values of people evolve as a linear function of their own and their neighbors' previous opinions. i.e.,

$$x_{k+1}^i = \sum_{j=1}^{|V|} A_{i,j} x_k^j, \ \forall k = 1 \dots K \tag{1}$$

Here, k represents a discretized time index. We will elaborate more on the time indices in Section 4, where we describe various modeling scenarios. $A_{i,j}$ represents the weight or intensity with which formation of node i's opinion at time $(k+1)$, x_{k+1}^i, gets influenced by node j's opinion x_k^j. The following constraint is naturally imposed on weights, since node j cannot influence node i if they are not connected:

$$(i,j) \notin E \quad \Longrightarrow \quad A_{i,j} = 0.$$

Also, $A_{i,i} \neq 0$ represents the weight with which agent i influences itself. Thus $\mathbf{A}$ can be thought as a weighted adjacency matrix of the graph G with all the self edges present. Let $\mathbf{x}_k = [x_k^1, \dots, x_k^N]$ denote the vector of all opinions at time k. We have the following equation representing the opinion dynamics:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k \tag{2}$$

Note that for $(i,j) \in E$, $A_{i,j}$ can be either positive or negative. A negative $A_{i,j}$ implies that agent/node i does get influenced by j's opinion, but to the opposite polarity. As a common example from real life, person i may know that her taste in movies is the opposite of person j. Hence, person j liking a movie may negatively influence person i's opinion about it. This effect is not possible in DeGroot's model [7], since $A_{i,j}$s are restricted to be positive and sum to 1. On the other hand, this assumption keeps the opinions predicted by DeGroot's model at time $k+1$ in the same range as the opinions in time k, thus imposing the bounds. The opinions predicted by the proposed model do not have a fixed bound. However, it is easy to check that:

$$\|\mathbf{x}_{k+1}\| \leq \|\mathbf{A}\|\|\mathbf{x}_k\| \leq \sqrt{\lambda_{max}(\mathbf{A}^T\mathbf{A})}\|\mathbf{x}_k\|$$

where, $\lambda_{max}(\mathbf{A}^T\mathbf{A})$ is the largest eigenvalue of $\mathbf{A}^T\mathbf{A}$. Hence, $\sqrt{\lambda_{max}(\mathbf{A}^T\mathbf{A})}$ imposes a dynamic bound on the predicted opinions.

Another aspect of our study is that we focus on short-term or bounded-horizon opinion dynamics, as opposed to asymptotic behaviour of an opinion dynamics model. Therefore, we can allow the use of models for which $\sqrt{\lambda_{max}(\mathbf{A}^T\mathbf{A})} \neq 1$. In the familiar asymptotic scenario, $\sqrt{\lambda_{max}(\mathbf{A}^T\mathbf{A})} > 1$ leads to divergence of opinions, while all opinions shrink to 0 if $\sqrt{\lambda_{max}(\mathbf{A}^T\mathbf{A})} < 1$. The focus on short term dynamics is fueled by the thought that influence of a person j on a person i, A_{ij} changes with time. In the experiments, we try to predict the opinions of $(k+1)^{th}$ timepoint using opinions of previous k timepoints. Next we describe metrics for evaluating the quality of predictions using data from social networks.

3.2 Metrics

In this paper, we adopt a data-driven approach to opinion modeling. To this end, we assume that we have access to actual opinions (ground truth) expressed by people interacting on social network (see Section 5). Given an algorithm that learns edge influence parameters, and a data set with ground truth opinion values at some time steps, we need metrics by which to evaluate the algorithm.

3.2.1 Normalized error

As before, let $x_k^i \in \mathbb{R}, i = 1, \dots, N$ be the opinion values expressed by users at timepoints $k = 1, \dots, K$. For real opin-

ions, a natural measure of error is the squared error of the predicted opinion with respect to the observed opinion. Thus error, $e^i_{k+1} = |x^i_{k+1} - \sum_{j=1}^N A_{ij}x^j_k|$. Hence, the root mean square error for all nodes at time $k+1$ is given by:

$$E_{k+1} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(e^i_{k+1})^2)} = \sqrt{\frac{1}{N}\|\mathbf{x}_{k+1} - \mathbf{A}\mathbf{x}_k\|^2}$$

However, this error metric is sensitive to the scale of the input data. Hence we use the *normalized error* metric:

$$NE_{k+1} = \frac{E_{k+1}}{(x_{max} - x_{min})} \tag{3}$$

where, $x_{max} = \max(x^i_k), \forall(i)_{i=1}^N$ & $\forall(k)_{k=1}^K$, and $x_{min} = \min(x^i_k)$, $\forall(i)_{i=1}^N$ & $\forall(k)_{k=1}^K$, are the maximum and minimum values of all observed opinions, respectively.

3.2.2 Quantized error

Another metric which captures the polarity of the opinions is the quantized error. We define this as the fraction of times, the polarity of the predicted opinion matches the observed one. Thus the quantized error at time instant $k+1$ is given by:

$$QE_{k+1} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left[x^i_{k+1}\sum_{j=1}^{N}A_{ij}x^j_k < 0\right] \tag{4}$$

where $\mathbf{1}(\cdot)$ is the indicator function. The product $x^i_{k+1}\sum_{j=1}^N A_{ij}x^j_k$ is positive only if x^i_{k+1} and $\sum_{j=1}^N A_{ij}x^j_k$ have the same sign.

4. LEARNING OPINION DYNAMICS

The model proposed above is specified by the set of parameters, $A_{i,j},\ i,j \in \{1,\ldots,|V|\}$. For a practical social network, it is difficult to ascertain $A_{i,j}$s manually. However, as described in Section 5, it is possible to obtain the opinions of various agents x^i_k at different time instants (see Section 5). Thus, the problem of automatically learning $A_{i,j}$-s given x^i_k-s is of importance. In this section, we explore various scenarios in which opinion data can be acquired and then describe methods of learning the model parameters under such scenarios. The next section (4.1) describes the scenario where every expressed opinion is captured and processed, while Section 4.2 describes a scenario relevant to large social networks with a large volume of opinions being exchanged, where it is not possible to record every exchange.

4.1 Asynchronous opinion dynamics

The simplest approach to learning a linear model described in Section 3.1 is to use opinion extracted from the individual posts by the users. In this scenario, the opinions are posted asynchronously, i.e., at time k, if an agent (j) posts its opinion, another agent i may not post any opinion. Let S be the set of all time instants when an agent has posted its opinion. Moreover, let $S_i = \{k|x^i_k \text{ exists}\} \subseteq S,\ \forall i = 1,\ldots,N$, be the set of all time instants when agent i has expressed an opinion. Let x^i_{k-}, be the last posted opinion by an agent i before time k.

$$\begin{aligned} x^i_k &= A_{i,i}x^i_{k-} + \sum_{j\in N(i)} A_{i,j}x^j_{k-} \\ &= \mathbf{A}_i^T\mathbf{x}_{k-}\ , \forall k \in S^i \text{ and } 1 \le i \le |V| \end{aligned} \tag{5}$$

where $N(i)$ is the set of neighboring vertices of i, A_i is the column vector i^{th} row of A, and $\mathbf{x}_{k-}$ is the column vector of all x^i_{k-}s. We call this model the **asynchronous linear model** (AsLM).

We assume that agent i forms its opinion at time $k \in S_i$ based on previously posted opinions of its neighbors. Let $\mathcal{D} = \{x^i_k | k \in S_i, i \in V\}$ be a dataset of all opinions posted by all agents in V. Assuming that x^i_k are plagued by additive zero mean gaussian noise, the loss incurred in predicting all observations by agent i is given by $\sum_{k\in S_i}||x^i_k - \mathbf{A}_i^T\mathbf{x}_{k-}||^2$. Adding an L_2 regularizer, $\lambda||\mathbf{A_i}||^2$, we can estimate the optimal parameter $\mathbf{A}^*_i$ by solving the following problem:

$$\begin{aligned} &\min_{\mathbf{A_i}} \sum_{k\in S^i} ||x^i_k - \mathbf{A}_i^T\mathbf{x}_{k-}||^2 + \lambda||\mathbf{A}_i||^2 \\ &\text{s.t.} \quad A_{i,j} = 0, \text{ whenever } ((i,j) \notin E)\&(i \neq j)) \end{aligned} \tag{6}$$

Here, λ is the user defined regularization parameter and $A_{i,j}$ is the j^{th} entry of vector $\mathbf{A_i}$. By solving $|V|$ such optimization problems (one for each i), we can obtain $\mathbf{A^*_i}, i = \{1,\ldots,|V|\}$, and thus estimate the entire adjacency matrix $\mathbf{A}^*$.

Let $\tilde{x}_{k-} = I_{ij}\mathbf{x}_{k-}, \forall k \in S_i$, where I_{ij} is a $N \times N$ diagonal matrix such that $I_{ij}(j,j) = 1$ if $(i,j) \in E$. Also, let $\mathbf{X}^i = [\tilde{x}_{k-}|k \in S_i]^T$ be a $|S_i| \times N$ matrix with rows as $\tilde{x}_{k-}$, and $\bar{x}^i = [x^i_k | k \in S_i]^T$ is a $|S_i| \times 1$ column vector. The above problem is same as solving $\mathbf{A}^*_i = \arg\min_{\mathbf{A}_i}(\|\bar{x}^i - \mathbf{X}^i\mathbf{A}_i\| + \lambda\|\mathbf{A}_i\|^2)$. It is easy to check that this problem is solved when:

$$\mathbf{A}^*_i = ((\mathbf{X}^i)^T\mathbf{X}^i + \lambda\mathbf{I})^{-1}(\mathbf{X}^i)^T\tilde{x}^i \tag{7}$$

$||\mathbf{A}||_F = \sqrt{tr(A^TA)}$ is the Frobenius norm of $\mathbf{A}$, where $tr(\cdot)$ in the trace operator. It is clear that increasing λ decreases $\|\mathbf{A}^*\|_F$ which can be thought of as a measure of complexity of the model (see e.g. [23]).

4.2 Time-aggregated opinion dynamics

The previous section described an asynchronous opinion propagation model. However, modeling every opinion expressed by the agents in a real social network is both computationally expensive and vulnerable to noise. For example, a 1% sample of tweets results in $\sim$ 500 million tweets in a day, each potentially expressing an opinion. Moreover, opinion values extracted from each tweet may be noisy due to fluctuations in mood of the agent, reaction to a previous tweet, etc. A time average, say over successive windows of an hour or a day, of the opinions extracted from tweets of an agent is expected to be less severely affected by such problems. In this section, we describe the dynamics of time aggregated opinion in two possible scenarios: periodic and aperiodic.

4.2.1 Periodic synchronous opinion propagation

Following Section 3, let $G = (V,E)$ be a graph representing a social network, where E also includes all self edges. Also, let $y^i_k,\ i = 1,\ldots,|V|$ represent the time aggregated opinion of i^{th} agent in the k^{th} time window. For simplicity, we assume that all the opinions y^i_k are available. This assumption is acceptable because one can always choose a suitable aggregation interval such that all agents have expressed at least one opinion in a time window. Under this assumption we can use the synchronous propagation model described in equation 2 to model the dynamics of opinion propagation:

$$\mathbf{y}_{k+1} = \mathbf{A}\mathbf{y}_k, \quad \forall k = 1,\ldots,K$$

where $\mathbf{y}_k$ is a vector of all time aggregated opinions, y^i_k, for the k^{th} time window. $\mathbf{A}$ is the weighted adjacency matrix governing the dynamics of time aggregated opinions. So, $A_{i,j} = 0$, if $((i,j) \notin E)\&(i \neq j)$.

A drawback of the above model is that it assumes opinions have been propagated only once during a time window (k to $k+1$). For

example, if the distance between i and j is greater than 1, there is no way for y_k^i to influence y_{k+1}^j. While this assumption was true for the scenario described in section 5.1, where every opinion propagation was considered, it is no longer valid in the time aggregated opinions. This is because there can be multiple rounds of opinion passing in a single time window. It is common for the user not to post every opinion she forms, or the analyst not be able to get hold of data at all time stamps or a combination of both.

To address this drawback, we propose the *periodic synchronous* opinion propagation model. In this model we assume that opinions are propagated periodically, with a constant frequency, say t per time window. Thus, using simple calculations, we can write the propagation model as:

$$\mathbf{y}_{k+1} = \mathbf{A}^t \mathbf{y}_k \ \ \forall k = 1, \ldots, K \tag{8}$$

We refer to this model as the **periodic linear model** (PLM).

Following the assumptions laid out earlier in this section, we can write the regularized loss function for learning $\mathbf{A}$ as $L(\mathbf{A}) = \sum_{k=1}^{K} \|\mathbf{y}_{k+1} - \mathbf{A}^t \mathbf{y}_k\|^2 + \lambda \|\mathbf{A}^t\|^2$. The best estimate of $\mathbf{A}$ can be obtained by minimizing $L(\mathbf{A})$. Unfortunately, $L(\mathbf{A})$ is not convex in $\mathbf{A}$. Hence the minimization can get stuck in local minimum. Also, we note that for most prediction tasks, we only need to estimate $\mathbf{M}_t = \mathbf{A}^t$, since we only observe opinions $\mathbf{y}_k$ which are propagated with the constant frequency of t per time window.

Let $G^t = (V, E^t)$ be the graph generated by including all t-hop connections in the set of edges E^t. Note that since the original graph has self-loops, E^t contains edges which can be traversed in *at most* t *hops* in the original graph $G = (V, E)$. It is clear that $M_t(i, j) = 0$ if $(i, j) \notin E^t$. Hence, we can learn the optimal $\mathbf{M}_t^*$ by solving:

$$\min_{\mathbf{M}_t} \sum_{k=1}^{K} \|\mathbf{y}_{k+1} - \mathbf{M}_t \mathbf{y}_k\|^2 + \lambda \|\mathbf{M}_t\|^2 \tag{9}$$
$$\text{s.t.} \quad \mathbf{M}_t(i, j) = 0, \text{ whenever } (i, j) \notin E^t$$

One way of obtaining $\mathbf{A}^*$ from $\mathbf{M}_t^*$ is to calculate $\mathbf{A}^* = (\mathbf{M}_t^*)^{1/t}$ using one of the root finding algorithms [12]. However, the solution in not unique. In general, there can be up to t distinct roots $(\mathbf{M}_t)^{1/t}$, and there is no way of ascertaining which of them is the correct one, unless we have opinions for intermediate steps within a time window. Another problem is that, the roots $(\mathbf{M}_t)^{1/t}$ may be complex (i.e., with imaginary parts), thus making interpretation of entries of $\mathbf{A}$ difficult. In the next section, we describe the aperiodic synchronous update setting, where some of the above problems are addressed.

4.2.2 Aperiodic synchronous opinion propagation

As in the previous section, we assume that time aggregated opinions are propagated synchronously. However, we assume that the number of times opinions are propagated in each time aggregated step can vary from one time window to another. The main motivation behind this assumption is that human activities happen in bursts. For example, people post more messages on social network during the day, than at night. Hence, it is expected that opinions will propagate further during a 6-hour time aggregate during day than the same period during night.

As before, let $\mathbf{y}_k$ denote the opinion vector for all agents at time k. Let $t_k, k = 1, \ldots, K$ be the number of times opinion propagates during k^{th} time aggregate. The opinion dynamics is given by:

$$\mathbf{y}_{k+1} = \mathbf{A}^{t_k} \mathbf{y}_k, \forall k = 1, \ldots, K \tag{10}$$

Note that the model is characterized by parameters $t_k, k = 1, \cdots, K$, in addition to the weighted adjacency matrix parameters A. The set of parameters $\mathcal{T}_K = \{t_k | k = 1, \cdots, K\}$ is called the *skip set*, with t_k denoting the number of iterations which has been "skipped" at k^{th} time aggregate. We denote the above model as **aperiodic linear model** (ALM). Analogous to previous discussion, we can write the following optimization problem for learning the weighted adjacency matrix parameter using the squared error as:

$$\min_{A} \sum_{k} ||\mathbf{y}_{k+1} - A^{t_k} \mathbf{y}_k||^2 + \lambda \|A\|_F^2 \tag{11}$$
$$\text{s.t:} \quad A_{i,j} = 0, \quad \forall (((i, j) \notin E) \& (i \neq j))$$

Note that here we assume the skip set $\mathcal{T}_K$ to be given. In practice, we can restrict each t_k to take values from a set $\{1, \cdots, t_{max}\}$. A search over all possible values of skip set will need search over $O(t_{max}^K)$ combinations. In section 6.3, we study some heuristic methods for fixing $\mathcal{T}_K$.

The above optimization problem is an instance of non-convex optimization problem in the matrix variable A. We find a local optimum for the above problem using projected gradient descent method, since the feasible set is convex. Let $\mathcal{Y}_K = \{\mathbf{y}_k | k = 1, \cdots, K\}$ be the set of all opinions. Let $f(A; \mathcal{Y}_K, \mathcal{T}_K) = \sum_i ||\mathbf{y}_{k+1} - A^{t_k} \mathbf{y}_k||^2 + \lambda \|A\|_F^2$. The gradient of $f(A)$ w.r.t. A can be written as

$$\nabla_A f(A; \mathcal{Y}_K, \mathcal{T}_K) = \sum_i t_k \big[-2A^{t_k - 1} \mathbf{y}_k \mathbf{y}_{k+1}^T + \mathbf{y}_k \mathbf{y}_k^T (A^{t_k})^T A^{t_k - 1} + A^{t_k - 1} \mathbf{y}_k \mathbf{y}_k^T (A^{t_k})^T \big] + 2\lambda A \tag{12}$$

The projected gradient descent algorithm for finding optimal A is described in Algorithm 1. Here, the gradient matrix $\nabla_A f(A; \mathcal{Y}_K, \mathcal{T}_K)$ is evaluated using expression in equation 12. $LineSearch$ is a function which ensures that the function value satisfies sufficient descent condition after moving a step length s in the gradient direction. The projection step $\Pi(A, E)$ ensures that resulting A is projected back to the feasible set, i.e. $A_{ij} = 0$ if $(i, j) \notin E$. While in general the algorithm is not guaranteed to converge, in practice we see that it converges for all inputs.

Algorithm 1: Learning A using projected gradient descent.

Data: $G = (V, E)$.
Input : Opinion-vectors: $\mathcal{Y}_K$, Skip-set $\mathcal{T}_K$, Starting point A_0, Convergence threshold ϵ, Edge set E
Output: Weighted-adjacency matrix: A
initialize: $A \leftarrow \gamma A_0$
while $(||\nabla_A f(A; \mathcal{Y}_K, \mathcal{T}_K)|| \geq \epsilon)$ **do**
  $s \leftarrow LineSearch(f, A, \mathcal{Y}_K, \mathcal{T}_K)$;
  $A \leftarrow A - s \nabla_A f(A; \mathcal{Y}_K, \mathcal{T}_K)$;
  $A \leftarrow \Pi(A, E)$
Return A

5. DATA SETS

We collected five data sets to evaluate our algorithm, which can be made available for further research. In each case, we needed the network topology, and opinion values of a set of users over a period of time. The five data sets, sketched in Table 1, can be placed in two groups. The first three were generated by us, in-house, through carefully controlled and monitored social influence process. The last two are derived from Reddit and Twitter data, provided as-is. The distinction is that in the first three cases, we could read agent opinion values at the time granularity of our choice, so as not to miss any updates; whereas for the last two, we have no such control.

Dataset	# Nodes	# Edges	# Docs	Max Docs./Node	Min Docs./Node
Continents: Europe vs North America	102	1,020	2,182	52	6
Colleges: IIT Delhi vs IIT Bombay	102	1,020	1,758	40	3
Occupation: Startup vs Job	102	1,020	1,439	33	4
Reddit (politics network)	556	94,312	64,366	2,571	20
Twitter (elections)	548	5,271	20,026	102	20

Table 1: Summary of the five data sets used for experimental validation. The first three correspond to the topics used for in-house controlled experiments on human subjects. The last two correspond to real world data sets obtained from Twitter and Reddit.

The first three cases provide us with valuable insight, as in these cases we were able to capture all visible opinion values, while also minimizing the influence of external sources.

5.1 Controlled social experiments

The set of agents in our controlled experiment was a class of 100 students in Information Retrieval course in Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur. The experiment was performed in a laboratory setting, where each student sat in front of a terminal and interacted with ten other randomly assigned students (designated social neighbors) through a Web interface (as shown in Figure 1) for a period of one hour. [1]

The agents refined their opinions continuously by communicating with their neighbors using the text box. To avoid externalities, participants were not allowed to access the Web, or discuss anything with each other verbally. All communication through the interface was recorded. Social neighbors were kept anonymous, so that the agents did not get biased by the real-life identity of another agent.

To collect one dataset, we started by broadcasting to agents a **topic**, posed in the form of a comparison between two entities A and B. The three topics given to the students were; compare two places to live (**Continents:** Europe vs North America), two colleges to attend (**Colleges:** IIT, Delhi vs IIT, Bombay), and choices of occupation forming a startup or working at an established company (**Occupation:** Startup vs Job). These topics were chosen as most agents did not have a strong prior opinion, but had some knowledge about the subject. This was done to ensure that at least some of the agents would show changes in their opinion during the experiment. Every time an agent posted a message, the interface automatically reported the current opinion value, which was modeled as a real number in the range $[-1, 1]$. The sign represents polarity of the opinion (e.g. if joining a startup is preferred then the opinion score assigned tends to -1, while reverse is true for the alternate case), and the magnitude represents the degree of conviction. Only the message from an agent, and not his/her current quantitative opinion, is shown to neighboring agents. Agents were asked to make opinion messages self contained. Every experiment proceeded for one hour, after which the experiment was terminated. At the end of a live experiment, we obtain one dataset, containing every visible timestamped opinion of every agent.

5.2 Twitter Dataset on Delhi Assembly elections 2013

The Delhi Legislative Assembly elections of 2013 was a keenly contested event with three major parties (two old and one newly formed) winning roughly equal vote share. Hence there was a need for post-poll alliance, this triggered a huge discussion on social media with people enthusiastically expressing a lot of opinions. This provided us with a very good opportunity to measure the performance of our system. For testing our system, we used the Twitter search API to collect tweets containing the following hashtags: #BJP, #APP, #Congress, and #Polls2013. The first three of these represent the hashtags for the three major parties competing in the elections, while the fourth was the most popular hashtag corresponding to the event. We gathered tweets during the period of 9th to 15th December 2013. This period corresponds to the week following the declaration of results on 8th December 2013.

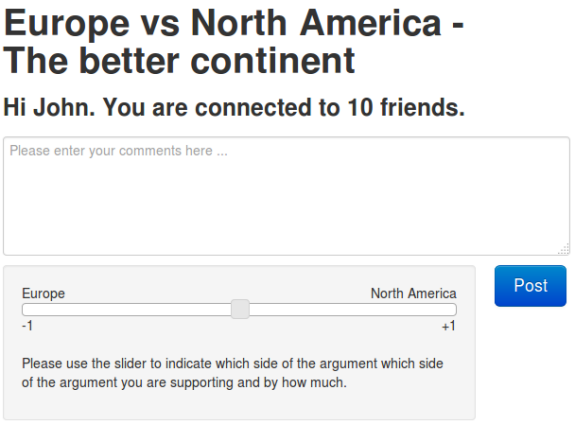

Figure 1: Web interface for opinion posting for the controlled experiment

5.2.1 *The network*

We filtered down to the candidate set of agents in three steps. We started with about 905,000 tweets, posted by 201,000 users. To remove corporate accounts, bot accounts, and spammers, we filtered the set of users based on the number of followees, number of followers, and the number of tweets posted by the user. We only preserved those users who had between 100 and 10,000 friends, between 50 and 1,000 followers, and between 200 and 10,000 tweets posted during the account's lifetime. This resulted in a set of 55,000 users. For these users, we collected the user IDs of all their followees, followers, and up to 3,200 most recent tweets using the Twitter REST API. We only collected tweets posted during the chosen week. With the information about both the followees and followers of the 55,000 users, we were able to create the complete follow network. Finally, from the 55,000 users, we selected the largest SCC such that each selected user has posted over 20 tweets. Thus we ended up selecting 548 users.

5.2.2 *Opinion values*

Since tweets are limited to only 140 characters, we grouped the tweets posted by every user during a single hour into a document. Each document is turned into an opinion score. 'Opinion' here con-

[1] In order to maintain both connectivity as well as randomness of the social graph with so few number of nodes, a realistic degree distribution like power-law could not be considered.

notes a positive or negative attitude to the political developments after election, which was detected by subjecting these hourly documents to a sentiment analysis tool specifically designed for Twitter [9]. For every document we finally get a single sentiment score in the range $[-1, 1]$. The score represents the relative proportions of words with positive and negative connotations.

Section 6 describes how the hourly sentiment scored documents for the 548 users and their follower network was used to understand the flow of opinion using our algorithm.

5.3 Reddit politics data

Reddit is a social post curation website, where users submit content in form of text posts or links to websites containing the content. More than 6% of online adult users use Reddit[2]. Content in Reddit is categorized by areas of interest called 'subreddits'. Reddit boasts over seven thousand active subreddits[3] on topics as varied as music, politics, sports, worldnews, programming, etc.

We collected data of Reddit users who posted content in the subreddit 'politics' during the period of July 1, 2012 to December 31, 2012. We crawled all posts made by Reddit users during the above period in the subreddit politics. We obtained 120,141 posts made by 30,812 users.

5.3.1 The network

The social network in Reddit is not explicit. We applied certain heuristics to recover and approximate the user network. We created an undirected network taking these $\sim$31,000 users as vertices, and assumed the existence of an edge between two users if there existed two subreddits (other than politics) where both posted during the given time period.

Similar to the case of the Twitter data, we randomly selected approximately 500 users such that the users have made more than 20 submissions during the given period and the network between them forms a single connected component. We ended up selecting a subnet of 556 users for the subsequent experiments.

5.3.2 Opinion values

Most of the posts made by users of Reddit are in well formed English. We used the standard linguistic analysis tool LIWC [20] to analyze sentiment scores from them. We computed the sentiment of a post as the difference between the positive emotion score and the negative emotion score, as returned by LIWC. The results were normalized by mapping the range of values obtained to the range $[-1, 1]$ using linear scaling.

Section 6 describes in more detail how the sentiment scored posts from the users were used to understand opinion propagation in Reddit.

6. EXPERIMENTAL RESULTS

In this section, we study and validate the models and methods proposed in this paper in a data driven manner. We report experimental results on the five datasets: for three of the dataset which have been produced at control setting and where each and every opinion of the participants are recorded, we perform experiments assuming asynchronous scenario and for two of the social network based data set, we perform experiment assuming synchronous scenario. Three techniques proposed here: one for the asynchronous setting (AsLM) and two for Synchronous setting (ALM and PLM). Four baseline models, *voter's model* [5], *biased voter's model* [6], *flocking model* [10, 6] and *DeGroot's model* [7]. To the best of our knowledge, this is the first work reporting a data-driven comparison of existing and new models for opinion dynamics using real-world as well as experimental datasets. Section 6.1 compares the performance of proposed models and learning methods with the baselines using metrics defined in section 3.2. Sections 6.2 and 6.3 describe techniques for choosing best period (periodic case) and skip set (aperiodic case), respectively. Section 6.4, validates the learning algorithm and observes its generalization ability.

6.1 Performance Comparison

For each of the models we perform a large number of experiments to chose the parameters so as to obtain the best result. In case of DeGroot model, we learn the parameters similar to our model. Side by side, we perform experiments on the three variations of the proposed scheme namely (a). AsLM (Asynchronous Linear Model) - here we consider that every opinion of each user is known, and the update of opinion can be different for different users. (b). PLM (Periodic Linear Model) - here we consider that the opinions are always updated synchronously after every (say) t time interval. (c). ALM (Asynchronous Linear Model) - here we consider that the time interval between any two opinion update varies. Since with the variation of time intervals, both ALM and PLM produce a lot of results, we report the best result here.

Table 2 reports a comparative analysis of the prediction-error for (AsLM/ALM/PLM) and the four state of the art algorithms - the first six(five) columns report the normalized mean square error (actual opinion prediction error) while the rest report the quantized error (polarity prediction error). We observe that for these two datasets, the overall performance of our schemes is substantially better than all the baselines. Out of the baseline, we see that ALM is performing better than PLM which confirms our proposition that users submit opinions at arbitrary instances.

Performance Analysis - Normalized RMSE: The first six (five) columns of Table 2 give a comparative view of actual opinion-prediction error.

Voter Model: Performance of Voter model is particularly poor. It relies on random opinion updates, thus evidently loses information of actual dynamics. Moreover such versions of voter model keeps the set of opinions in a graph invariant throughout the process. This intrinsic property of voter model prevents the opinion-values from not growing in a larger space which thereby goes against the spirit of continuous opinion-model. Biased Voter Model attempts to overcome these by introducing node weights. However the performance of Biased Voter model is worse than ALM or PLM. A closer scrutiny reveals that, biased voter model parameterizes the node weights; however due to uniform edge weights, it is unable to capture the actual influence dynamics.

Flocking Model: Note that the RMSE for flocking is substantially lower than other three baselines in most cases. But it performs poorly in predicting the polarity which is reflected in relatively higher values of quantized error. This is because flocking model assigns more weights to opinions "close" to user's own opinion. In this model the "closeness" solely depends on the absolute difference in opinion-values and hence the polarity difference is neglected.

DeGroot Model: The performance of DeGroot model is fairly competitive for Reddit and Twitter. This is mainly because it incorporates different edge weights that capture the actual dynamics of information-flow from one node to another, which is heavily neglected in the other three baselines. The relatively better performances of flocking and DeGroot model also reflects an inherent linearity in the dynamics that justifies our choice of a more generic

[2] `http://pewinternet.org/Reports/2013/reddit.aspx`

[3] `http://www.reddit.com/about/`, as on June 7, 2014.

linear model.
ALM and PLM perform significantly better than all the baselines. A possible explanation can be that it captures the effect of intermittent observations i.e. the phenomenon of periodic/aperiodic observations, which neither of the baseline-algorithms takes care of. Our model is also not limited to positive entries and row-stochasticity, which are the major features of DeGroot model. Being the most generic linear model it captures the negative influence, opinion fluctuation etc. It also allows formation of any generic linear combinations of opinions rather than convex combinations.

Variation across the Datasets: From table 2 we observe that the algorithms perform substantially better in Reddit than in Twitter. Note that in case of Reddit we have collected the evolution of general political opinion whereas in Twitter we concentrated on a specific event, corresponding to a legislative election. Reddit is a forum, where people actually join to form an opinion/impression. Therefore, it is natural that a user in Reddit view others' post, form an opinion and write a well-thought post. Also since the users are more in exploratory mode, a Reddit user can read and scrutinize any other people's comments, which evidently helps her to form an opinion. In our model we have taken a decent estimate whereby two agents are neighbors if they have subscribed to three common sub-reddits, even then we find that the reach of each agent is a magnitude higher than that of Twitter.

On the other hand Twitter is a popular social-network site and we are looking into the data of a particular popular event. Since the underlying graph structure is sparse, an opinion may take time to propagate and may get lost in the process [15]. Thus the effect of a distant node becomes almost negligible. Also since the event tracked is popular, much of the information may be coming from (outside) Twitter and a user's opinion may get influenced due to that [18]. Therefore, PLM/ALM which assume local influence perform worse in capturing the influence dynamics.

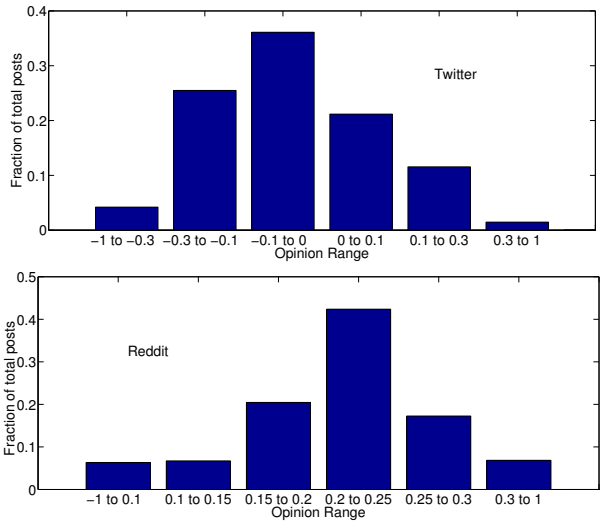

Figure 2: Distribution of Opinion-Values in Datasets

Performance Analysis - Quantized Error: It is seen that the quantized error for Reddit is even smaller than Twitter, this happens because in Reddit the opinion is in general positive so the chance of making sign error diminishes while in Twitter, since we are considering opinion pertaining to a particular event, there is a healthy mix of positive and negative opinions and the mean of the opinion is around zero, so the vulnerability of the dataset is much higher.

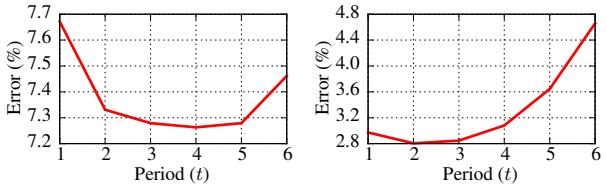

Figure 3: Error(%) vs Period(t) for (a).Twitter and (b).Reddit datasets.

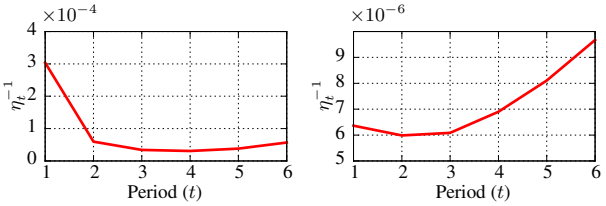

Figure 4: η_t^{-1} vs Period(t) for (a).Twitter and (b).Reddit datasets.

From figure 2, we observe that less than 10% posts in Reddit are negatively polarized. In such cases, analysis of continuous-opinion is important as a huge portion of the users are biased towards particular side. For the three social games, the performances of all the algorithms are substantially better. Our model gives a 100% accuracy in these three games. This is because the active and enthusiastic participation of the users in the experiments lead to a uniformly polarized dataset of opinions with a nice dynamical flow. Hence all the algorithms are able capture the dynamics of the process with high prediction accuracy.

6.2 Synchronous - Periodic

Figure 3 describes the variation of normalized error (NE) with respect to a given period t. From the figure it is clear that in both the cases, the minimum error is reached at an interval value greater than 1 (4 for Twitter and 2 for Reddit). The best period t that minimizes error gives a good estimate of the typical number of opinions posted by a user between two successive opinion aggregations as part of our data processing steps. This hypothesis can be corroborated by defining the following two quantities, and visualizing their relation with the period t, as compared with error vs. t. This study is shown in Figures 4 and 5.

Effective Paths (η_t) - It measures the number of pairs which has more than t number of (weighted) t-hop paths. It can be represented as

$$\eta_t = \sum_{i,j} \frac{\mathbf{1}(P_t(i,j) \geq t)}{2^t} \tag{13}$$

where $P_t(i,j)$ is the number of $t-hop$ paths between a pair (i,j), $\mathbf{1}(\dot{})$ is an indicator function. A forgetting factor (2^t) to effectively model the increase in hop distance is added.

t-opinion heterogeneity $\mathcal{E}_t$ - Here the idea is that if an agent influences nodes at t hop distance, then its t-hop neighbors would have similar value. The heterogeneity in value can be measured using

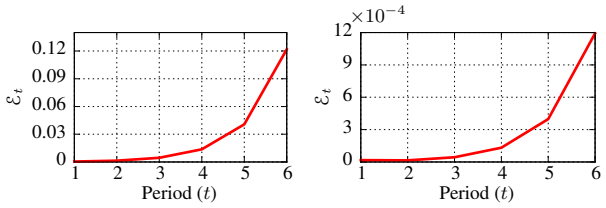

Figure 5: $\mathcal{E}_t$ vs Period(t) for (a).Twitter and (b).Reddit datasets.

Dataset	Normalized RMSE						Quantized Error					
Synchronous scenario (Periodic and Aperiodic)												
	PLM	ALM	BiasedVoter	Voter	DeGroot	Flocking	PLM	ALM	BiasedVoter	Voter	DeGroot	Flocking
Twitter	7.26	7.16	17.49	22.98	10.20	9.49	7.85	7.84	14.78	15.33	12.96	24.09
Reddit	2.80	2.75	7.51	15.60	6.00	8.24	1.08	1.08	2.16	2.70	1.62	10
Asynchronous scenario												
	AsLM		BiasedVoter	Voter	DeGroot	Flocking	AsLM		BiasedVoter	Voter	DeGroot	Flocking
Continents	10.42		31.46	35.51	23.94	32.89	0		1.96	2.94	1.96	5.88
Colleges	12.80		22.77	28.69	59.28	32.06	0		2.94	3.92	2.94	4.90
Occupation	10.36		23.06	30.32	33.28	31.64	0		2.94	6.86	0.98	7.84

Table 2: Comparison of performance between 3 models proposed here and 4 baseline models using 5 datasets and 2 error metrics. (See section 6.1)

the concept of Shannon entropy. So to calculate $\mathcal{E}_t$ we measure the opinion of the set comprising of the node and its t-distant neighbors. It can be represented as

$$\mathcal{E}_t = \frac{1}{NK}\sum_{i=1}^{N}\sum_{\tau=1}^{K}\sigma(\Psi_t(\tau,i)). \tag{14}$$

where $\sigma(.)$ is the standard deviation of any vector-set $(.)$, i.e. $\sigma(S) = \sqrt{\frac{1}{|S|}\sum_{i=1}^{|S|}(S(i)-\bar{S})^2}$. $\Psi_t(\tau,i) = \cup_{j\in V}\{y_j(\tau)|(i,j)\in E^t\}$, $y_j(\tau)$ is the opinion of j at time τ, N is the total no. of nodes and K is the maximum time-stamp of opinions in the dataset.

In the following, we investigate the behavior of Twitter and Reddit individually.

Reddit: The second part of Figure 4 plots the variations of period t with η_t^{-1}. It is seen here that the number of effective paths reaches an optimal point around $t = 2$; interestingly the curve is exactly similar to the second part of Figure 3 which corresponds to the Twitter dataset. Side by side, when we measure the t-opinion diversity, we see that it quickly diverges beyond 3. So 2-3 is the effective hop where we find the influence of an agent is maximum after one post.

Twitter: In case of Twitter, the first part of Figure 4 plotting the variations of period t vis-a-vis η_t^{-1} shows that at hop distance 4, the most number of effective paths are found which explains the value of 4. Also check the figure 5a, the 4-opinion-heterogeneity is significantly low.

6.3 Synchronous Aperiodic Data

In this section, our objective is to heuristically find the best choice of skip-set $\mathcal{T}_K = \{t_k|k = 1,\ldots,K\}$. For this experiment, we randomly generate 500 skip-sets such that $t_k \in \{1,\ldots,6\}$. For each of these, we compute the optimal A using algorithm 1, and calculate the resulting errors. The figures in this section explore correlation between these errors and some observed properties of skip sets.

Variation of error w.r.t $\sigma_{\mathcal{T}_k}$: Let $\sigma_{\mathcal{T}_k} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(t_k-\mu)^2}$, where $\mu = \frac{1}{K}\sum_{k=1}^{K}t_k$, be the standard deviation of skip-set $\mathcal{T}_k$, measuring the homogeneity of $\mathcal{T}_k$. Hence $\sigma_{\mathcal{T}_k} = 0$ corresponds to the periodic case. From figure 6 we see that $\sigma_{\mathcal{T}_k}$ is correlated with error in a continuous manner, with the function reaching a minimum at a point different from 0 in both Twitter and Reddit datasets. Hence, the effect of bursty nature of human activity in these datasets is evident. Moreover the figure suggests that deviation from the optimal periodic skip, t, is an important parameter, leading to the next study.

Variation of error w.r.t $\Delta_{\mathcal{T}_k}$: We define $\Delta_{\mathcal{T}_k} = \frac{1}{K}\sum_i(k_i - k_{\text{per}}^{\text{best}})^2$ as the deviation from best periodic skip obtained in section 6.2. Hence for Twitter $k_{\text{per}}^{\text{best}} = 2$ and for Reddit $k_{\text{per}}^{\text{best}} = 4$. Figure 7 shows a similar variation of error w.r.t $\Delta_{\mathcal{T}_k}$. We observe that the best accuracy of prediction is obtained at a non zero value of $\Delta_{\mathcal{T}_k}$. This evidently supports the utility of aperiodic model in opinion propagation. We observe that in case of Reddit, the aperiodic skip giving minimum error is much further from periodic compared to Twitter. This might suggest a more burst nature of activity in Reddit than Twitter.

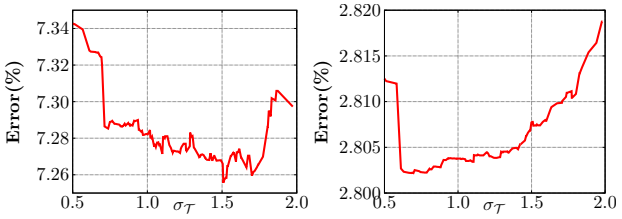

Figure 6: Error(%) vs $\sigma_{\mathcal{T}}$ for Twitter and Reddit datasets.

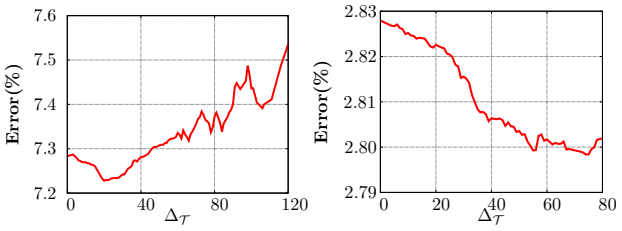

Figure 7: Error(%) vs $\Delta_{\mathcal{T}}$ for Twitter and Reddit datasets.

6.4 Validation of Learning Parameters

Table 3 shows variation of training set error ($\epsilon_{\textbf{TrainSet}}$), test error ($\epsilon_{\textbf{TestSet}}$) and the measure of model complexity ($\textbf{Tr}(A^TA)$) with the regularization parameter λ. We note that at low values of λ the training error $\epsilon_{\textbf{TrainSet}}$ is relatively smaller for all datasets. But in such cases, the test-set error ($\epsilon_{\textbf{TestSet}}$) is high. This shows that at low values of regularizer (λ) the learning causes overfitting the parameters. It is also evident from large values of $\text{Tr}(A^TA)$ which in turns means that the complexity of model is too high. However as the λ value is increasing the linear model attempts to capture the dynamics more and more efficiently and the $\epsilon_{\textbf{TestSet}}$ goes down. Large value of λ penalizes the regularized loss too heavily and the resultant learning-performance becomes very poor. So there exists an optimal regularizer value (λ_{opt}) that gives a best-fit for the model. By tuning λ, one can obtain the optimal value of λ, that results in the least prediction error $\epsilon_{\textbf{TrainSet}}$ and corresponding the most effective regularization-parameter (λ_{opt}) for the model.

Topic/Dataset	λ	$\epsilon_{\textbf{TrainSet}}$	$\epsilon_{\textbf{TestSet}}$	$\text{Tr}(A^T A)$
Colleges	0.2	0.0109	0.147	62.09
	0.5	0.0116	0.135	41.11
	1.5	0.0129	**0.128**	26.61
	4.0	0.0147	0.134	18.34
	10.0	0.0178	0.157	12.11
Continents	0.7	0.0113	0.109	18.07
	1.4	0.0119	0.105	12.75
	2.8	0.0127	**0.104**	9.02
	5.6	0.0136	0.106	6.42
	11.2	0.0147	0.110	4.75
Occupation	0.3	0.0095	0.114	37.45
	0.6	0.010	0.1139	37.45
	1.3	0.0109	**0.103**	19.51
	2.6	0.0513	0.1053	14.315
	5.2	0.0134	0.1169	9.99

Table 3: Efficacy of Learning (Sec. 6.4) - Variation of errors $\epsilon_{\textbf{TrainSet}}$ and $\epsilon_{\textbf{TestSet}}$ and model complexity ($\text{Tr}(A^T A)$) w.r.t regularizer (λ). λ_{opt} is indicated in bold font.

7. CONCLUSION

We presented a family of algorithms for estimating edge (social link) influence strength in a social network from observing the state of quantitative opinions evolving along time at the nodes (representing actors or agents). Unlike some earlier work on continuous opinion dynamics, we do not seek or depend on asymptotic or steady-state behavior. We also presented variations where opinion data is pre-aggregated, and/or we cannot observe the global state of opinions at all times. Experiments with five data sets showed that our estimates of edge influence strength let us estimate node opinions much more accurately than several baseline models for influence propagation. Other contributions include proposed metrics to evaluate such influence estimation algorithms, along with three real-life data sets that we created.

Acknowledgement: This work was partially supported by Google India under the Google India PhD Fellowship Award, a fellowship grant from Tata Consultancy Services and a grant from ITRA sponsored project "DISARM". We also thank Yahoo! for their travel support and donation of servers to Dept. of Computer Science and Engg., IIT Kharagpur.

8. REFERENCES

[1] R. Axelrod. The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, 41(2):203–226, 1997.

[2] D. Bindel, J. Kleinberg, and S. Oren. How bad is forming your own opinion? In *FOCS Conference*, pages 57–66, 2011.

[3] B. Chazelle. Natural algorithms and influence systems. *Commun. ACM*, 55(12):101–110, Dec. 2012.

[4] F. Chierichetti, J. Kleinberg, and S. Oren. On discrete preferences and coordination. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, pages 233–250, 2013.

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

[6] A. Das, S. Gollapudi, and K. Munagala. Modeling opinion dynamics in social networks. In *ACM Conference on Web Search and Data Mining 2014(to be published)*, pages 585–586.

[7] M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345), 1974.

[8] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *WSDM Conference*, pages 241–250, 2010.

[9] A. Hannak, E. Anderson, L. F. Barrett, S. Lehmann, A. Mislove, and M. Riedewald. Tweetin' in the Rain: Exploring societal-scale effects of weather on mood. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM '12)*, Dublin, Ireland, June 2012.

[10] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence: Models, analysis and simul ation. *Journal of Artificial Societies and Social Simulation*, 5:1–24, 2002.

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

[12] B. Iannazzo. On the newton method for the matrix pth root. *SIAM Journal on Matrix Analysis and Applications*, 28(2):503–523, 2006.

[13] D. Kempe, J. Kleinberg, S. Oren, and A. Slivkins. Selection and influence in cultural dynamics. In *ACM Conference on Electronic Commerce*, EC '13, pages 585–586, 2013.

[14] D. Kempe, J. Kleinberg, and E. Tardos. Influential nodes in a diffusion model for social networks. In *Proceedings of the 32Nd International Conference on Automata, Languages and Programming*, ICALP'05, pages 1127–1138, 2005.

[15] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a Social Network or a News Media? In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 591–600, New York, NY, USA, 2010. ACM.

[16] N. Lanchier. Opinion dynamics with confidence threshold: an alternative to the Axelrod model. 2010.

[17] J. Lorenz. Heterogeneous bounds of confidence: Meet, discuss and find consensus! *Complexity*, 15(4):43–52, 2010.

[18] S. A. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, pages 33–41, 2012.

[19] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, Jan. 2008.

[20] J. W. Pennebaker, M. E. Francis, and R. J. Booth. LIWC: Linguistic Inquiry and Word Count. `liwc.net`, 2007. Accessed on June 03, 2014.

[21] M. Richardson and P. Domingos. Mining the network value of customers,. In *SIGKDD Conference*, pages 57–66, 2001.

[22] S. Shahrampour, S. Rakhlin, and A. Jadbabaie. Online learning of dynamic parameters in social networks. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, editors, *NIPS Conference*, pages 2013–2021. Curran Associates, Inc., 2013.

[23] J. F. Trevor Hastie, Robert Tibshirani. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2001.

[24] E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione. Binary opinion dynamics with stubborn agents. *ACM Trans. Econ. Comput.*, 1(4), 2013.

[25] M. E. Yildiz, R. Pagliari, A. Ozdaglar, and A. Scaglione. Voting Models in Random Networks.