Tracking Individual Behaviors in Networks: An Experimental Demonstration*

Balakumar Balasingam  
Department of ECE  
University of Connecticut  
Storrs, CT 06269, USA  
bala@engr.uconn.edu

Peter Willett  
Department of ECE  
University of Connecticut  
Storrs, CT 06269, USA  
willett@engr.uconn.edu

Yaakov Bar-Shalom  
Department of ECE  
University of Connecticut  
Storrs, CT 06269, USA  
willett@engr.uconn.edu

Abstract—Tracking individual behaviors based on observations made from vast personal interaction networks has become a major concern and interest for the policing community as well as for the business/commercial players. While the policing community resort to personal networks in order to predict and prevent adverse events, the commercial players want to track opportunities for online advertisement, market identification, personalized product suggestions etc. Recently, revolutionary advances in digital media technology have enabled one to collect, store and analyze massive amounts of personal networking data. Unlike traditional tracking problems, the observations and the inferred targets are highly irregular in nature; they do not evolve or be observed according to established mathematical models.

In this paper, we demonstrate an experimental approach for tracking hidden qualities of individuals by observing their closer connections in the personal networks they belong to. We model the hidden features of individuals through hidden Markov random fields (HMRF) and propose a modified observation model in order to simplify the tracking algorithm. We test our algorithm on a fictitious scale-free personal network dataset and report high accuracy through objective performance metrics.

Index Terms—Dynamic networks, graph theory, latent states, non-traditional target tracking, hidden Markov random field (HMRF).

I. INTRODUCTION

Even though the concept of personal networks is as old as human civilization, the term has gained an important re-definition in the recent past. Almost a quarter of the world population is now connected through some kind of a personal media network which allows them to broadcast their opinion privately to their connections or in public to the rest of the entities in the network. As the majority of the population moves to the cyber networking, so has the business and governing system that historically existed in the physical world only. Unlike in the physical world, the cyber world does not have well defined borders or laws and orders. This poses opportunities as well as challenges in unique different ways. Recently there has been significant effort in understanding personal networking through mathematical modeling and simulations. In the rest of this section we review some relevant existing works that attempt to study dynamic personal network tracking.

In [1] a large scale dynamic network based on (AT&T residential) telephone calls consisting of several millions of nodes and billions of edges was studied. The primary objective was to analyze the dynamic network and detect malicious activities in the network. With rapid new entries and exits to the graph another important objective was to maintain a time-smoothed version of the graph at the current instant and predict the graph to a future time point.

Another important work in dynamic personal network tracking was the introduction of latent variables and the approaches to define and estimate them [2] [3], [4].

In [5], [6] dynamic social network theory and graph estimation approaches based on Markov random graph model or p* model was discussed. The use of Stochastic Ordinary Differential Equations (SODE) in dynamic social network analysis and tracking was discussed in [7] [8]. These approaches were applied to smaller social networks in the range of 1000 nodes. The use of these methods in large scale social networking applications is yet to be investigated.

Some of the past works focus on finding dynamic communities in social networks: The dynamic community finding and similarity detection algorithm [9], community evolution tracking method [10], the approach for matching clusters between time steps using similarity metrics [11] and the adaptive evolutionary clustering [12] are some examples.

Some other past works focused on modeling the evolution of dynamic social networks: Modeling the evolution of social networking [13], identifying and tracking dynamic processes from textual data [14], empirical analysis of an evolving social network [15], and a quantifying method for social group evolution [16] are some notable ones.

Some other works on dynamic network tracking include a discussion on what is trackable in social network domain [17], a Bayesian tracking approach of emerging epidemics using ensemble optimal statistical interpolation [18], a method for the estimation of “vertex-betwenness” using dynamic social information [19] and the approach for the extraction and mining of academic social networks [20].

So far, little effort has been made to model and estimate the time varying nature of attributes of an entity in the network. Time varying attributes try to model a personal behavior as realistically as possible and pose such questions as: How has

*Proc. 15th Inta’l Conf. on Information Fusion, Singapore, July 2012.
a person considered friendly at one time has evolved into adversary at another time; What external factors influence the product loyalty of a certain person? and Can we track such change in individual preference based on some observable data from the network?

First, we model the dynamic social network by accounting for the time varying attributes of the entities, nodes and edges, in the network. Then we experiment with probabilistic models to relate these temporal attributes to our fictitious ground truth. We then model the spatial as well as temporal dependencies in these variables through hidden Markov random field (MRF) theory. Finally, we propose a tracking algorithm that exploits the model in order to infer hidden features of persons in a dynamic network.

The rest of the paper is organized as follows: In Section II a dynamic network is defined in probabilistic domain. In Section III a recursive tracking algorithm for hidden features was derived. In Section IV the details of the experimental study is summarized and in Section V the paper is concluded.

II. MULTI-ATTRIBUTE DYNAMIC NETWORK

Let us consider a network consisting of \( N \) nodes. In a social network setting nodes shall refer to persons, places and events. There can be up to \( N(N - 1)/2 \) edges (or links) between these nodes. In a multi-attributed setting, each node as well as the edge will possess numerous attributes. For example, a node (let us say this particular node is a person) will have attributes such as: gender, age, ethnicity, occupation, language, hair color etc; A link will have attributes such as: type, how well connected, strength of link, frequency of communicative transactions etc.

Considering all of the above, a multi-attributed network of \( N \) nodes can be represented as a tensor of attributes, i.e.,

\[
X = \{x_{km}\} \quad k, m = 1, \ldots, N
\]

where

- The diagonal vectors \( x_{kk}, k = 1, \ldots, N \) are the attributes of the notes represented as a vector of \( L_{kk} \) attributes, i.e.,

\[
x_{kk} = [x_{kk}^1, \ldots, x_{kk}^{L_{kk}}]
\]

- The off-diagonal vectors \( x_{km}, k, m = 1, \ldots, N, \quad k \neq m \) are the attributes of the links, defined as a vector of \( L_{km} \) attributes, i.e.,

\[
x_{km} = [x_{km}^1, \ldots, x_{km}^{L_{km}}]
\]

and \( L_{km} \in \mathbb{J} \).

In a time varying context each of the network attributes defined above is assumed to have time varying values. Hence the time varying network attribute tensor is denoted as \( X(t) = \{x_{km}(t)\} = [x_{km}^1(t), \ldots, x_{km}^{L_{km}}(t)], \quad k, m = 1, \ldots, N \).

It is often convenient to decouple the above multi-attributed social network into different layers. We propose to do so by conceptually dividing the network into the following three layers

1) Latent states of interest: The ultimate objective of dynamic network tracking is to estimate and track latent states of the network. A latent state is usually subjective, e.g., the anti-US stance (or hostility) of a specific node in the network. The latent variables of interest \( L(t) \) is often related to the multi-attributed network, i.e., \( L(t) = f(X(t)) \). Latent states are not directly observable – any directly observable attribute is considered part of the true network described next.

2) The true network: When it comes to social networking, the idea of the true network is elusive. Some aspects of the true network have physical representations, e.g., the age of a person can be confirmed reasonably well. Some aspects are hard to represent in a network, e.g., an attribute of the link between a person A and a person named Jon Stewart. Even though real world representation of such attribute is elusive and difficult, that particular attribute might have a higher correlation towards the latent state (of person A) which describes as to which party the vote of A is for in the next election.

3) Observed network: The information regarding the nodes and links above can be collected/observed through different ways, such as, data mining, interviewing community members, etc.

Some insights into the above layers can be found in [2], [3], [4] as well. Please see Figure 1 for a graphical/intuitive illustration of the above three layers.

The observations are noisy and erroneous, e.g., the age of a person that is reported as 25 could actually be between 20–30; the eye color of a person that is reported as blue could be green. As a result, an observation model is required to accounts for errors in the collected data. In the most general form an observation model can be written as

\[
y_{km}(t) = h_{km}(x_{km}(t)).
\]

It should be noted that there is a high possibility of ambiguous observations. For example, according to data collected by interviewing people, a person A (node \( m \) in the network) is rich and has terrorist connections. In reality the person A indeed is rich and seems to be running a charitable orphanage. However, the reported data could be about person B (node \( n \) in the network) who is rich and does have suspicious connections. Hence it is important to incorporate this “data association” information within the observation model, i.e., the observation model should take into account that a rich person with terrorist connections could be person A with probability \( p_1 \) or another person with probability \( p_2 \) and so on. After processing a batch of data the tracking algorithm may be able to determine that the rich person with terrorist connections is indeed person B.

Let us assume that the observations are made at time instances \( t = 1, 2, \ldots, T \). Let us denote the observations at time \( t \) as \( Z(t) = \{z_{rk,m}(t)\}^{R_{rk,m}(t)}_{r_k=1, r_m=1} \). Let us also introduce parameters \( c_k(t), c_m(t) \)

\[
Pr(c_k(t)c_m(t) = \{km\}) = \pi_{km}
\]
such that

\[ z_{rkm}(t) = y_{ck(t)c_m(t)}(t) \]  

where \( \pi_{km} \) is a lower diagonal probability matrix.

The objective of the dynamic network tracking algorithm is now described as follows: Considering the collected observations \( Z^T = \{Z(1), \ldots, Z(T)\} \) determine the values for the node and link attributes \( X^T = \{X(1), \ldots, X(T)\} \) over the observation period.

The entire network \( X \) and the observations can be represented in the probability space \( (Z, X, \Pi(X|Z)) \) where \( \Pi \) is a probability density function defined by its arguments, i.e., \( \Pi(X|Z) = p_{X|Z}(X|Z) \).

The multi-attributed dynamic network \( X \) represents a random field, i.e., current quantities of each of the attributes are conditioned on all the other attributes at all times instances in the past, i.e.,

\[
X_{km}(t) = f(x_{ij}(t)) \quad \left\{ \begin{array}{l} i = 1, \ldots, N \\
 j = 1, \ldots, N \\
 t = 1, \ldots, t-1 \end{array} \right. \]  

(7)

It is often convenient to approximate the random field \( X \) via Markov properties, i.e., the dependency between random variables is assumed only limited to the neighborhood and dependency in time is only limited to previous scan. With this assumption, the state space in (7) is considered to be a hidden Markov random field (HMRF) and can be re-written as

\[ x_{km}(t) = f(x_{ij}(t))) \quad \left\{ \begin{array}{l} i, j \in N(k, m) \\
 t = t-1 \end{array} \right. \]  

(8)

where \( N(, \) denotes the neighborhood of the argument. Correspondingly, the observation is written as

\[ Z(t) = \{y_{k1}m1(t)(t), y_{k2}m2(t)(t), \ldots, y_{ks}m(t)(t)\} \]  

(9)
III. APPROXIMATION OF HMRF THROUGH MODIFIED OBSERVATION MODEL

The focus of this paper is limited to $L_{k,m} = 1$, i.e., we limit the number of possible attributes to be one. Hence, the HMRF assumption implies that the attribute of a person is only influenced by other people in the network that are in his neighborhood. In social network terms a neighborhood is measured in terms of hop-distance. To elaborate, if node $i$ and $j$ are directly linked then the hop distance from node $i$ to node $j$ is one; if $i$ and $j$ are not directly linked but they are both linked to another node $k$ then the hop distance from node $i$ to node $j$ is two.

At this point, we propose some intuitive approximations for the HMRF model of the network defined in (8). First, we reduce the scope of the tracking to single binary attributes of nodes only, i.e., we assume that only the nodes in the network have dynamic attributed that needs to be tracked and that the attribute of each node is either of a discrete binary value. Second, we approximate the attribute dependency of nodes in the neighborhood to the observations, i.e., in this case the HMRF $X$ is reduced to a Markov process as written below.

$$x_k(t) = f(x_k(t-1)) \quad (10)$$
$$y_k(t) = g(x_k(t)) \quad m \in \mathbb{N}(k) \quad (11)$$

Since we are only interested in tracking a single attribute of nodes (not links), from here on we use $S_i^t$ to denote the attribute of the node $i$ at time $t$, $x_i(t)$.

Let us assume that the attribute is discrete $S_i^t \in \{0, 1, N_x - 1\}$, i.e., the attribute is assumed to take discrete values. Now (11) can be expressed through the transition probability matrix $A$ as expressed below

$$
\begin{array}{c|ccc}
    & 1 & \ldots & N_x \\
\hline
1 & P_{1,1} & \ldots & P_{1,N_x} \\
\vdots & \vdots & \ddots & \vdots \\
N_x & P_{N_x,1} & \ldots & P_{N_x,N_x} \\
\end{array}
$$

(12)

where $P_{m,n}$ indicates the probability $p(S_{t+1}^i = n|S_t^i = m)$. The initial state $S_0$ is defined by a prior distribution $\pi = \{\pi_1, \ldots, \pi_N\}$. A graphical illustration of (12) for $N_x = 2$ is shown in Figure 2.

Based on our approximated HMRF model, the observations have to be made in the neighborhood of the $i^{th}$ node. Based on the structure of the experimental data we decided to use as observations the distance to a known or confirmed hostile node in the neighborhood of node $k$. Figure 3 details how this distance is measured.

By assuming the observation made for node $i$ at time $t$ to be $Y_i^t$, the observation likelihood probability $p(Y_i^t|S_i^t)$ can be summarized in the form of $N_x \times N_y$ matrix $B$ as illustrated in Figure 4. Here, $N_y$ denotes the number of possible observations. Now, based on the assumptions described so far, the HMRF representation of the dynamic attributes of the network $X$ is further decomposed into separate hidden Markov models (HMM) for each node $i$ in the network.

With no approximations, the HMRF model parameters can be estimated using the expectation maximization (EM) type algorithms [21]. For the modified HMRF model presented in this section the model parameter estimation reduces to estimating the HMM parameter $\lambda_i$ for each of the nodes. Assuming the availability of training data, this can be easily done through the Baum-Welch algorithm [22]. It must be noted that social network (SN) theory is an established field that can be a guide in the selection of model parameters. In the absence of training data SN theoretical avenues has to be pursued in order to arrive at reasonable models. Also, it is also possible that the attributes are defined through multiple models. In such events, modern data fusion techniques have to be explored [23]. Figure 5 illustrates a general model selection strategy in the event of uncertain knowledge regarding the underlying dynamics network attributes.

Remark: The algorithm in its current form is well scalable to

![Diagram](image-url)
linearly scales to larger networks. In independent observation models. As a result, the algorithm decoupled into binary state HMMs per each nodes with observations. The HMRF is essentially large networks due to the assumptions (approximations) made in the modified observation model. The HMRF is essentially decoupled into binary state HMMs per each nodes with independent observation models. As a result, the algorithm linearly scales to larger networks.

IV. EXPERIMENTAL STUDY

We based our study on a artificially simulated data set. The data consist of nodes representing individuals, groups and events, and their relationships to each other. Although the scale-free structure of the network is similar to that of any real-world social network, none of the actual nodes or relationships were based on or drawn from any real-world individuals or networks. The dataset included several hundred nodes some of which were described as hostile or non-hostile and a few thousand relationship links. The links were not weighted, hence, if person $k$ is linked to place $m$ then the link_exist property for the $(k,m)^{th}$ link is set to 1. A view of all the nodes and links in the network is shown in Figure 6.

Even though the above data represented a realistic social network it did not have any dynamic features or attributes. Hence we have used the following procedure to simulate a dynamic network: We assigned a portion of the entire nodes as “initial nodes” at time $t = 0$. All the links that have both of their ends among the initial nodes are declared as “initial links”. It is assumed that the initial nodes and links are known as prior values for the tracking algorithm. For $t = 1, 2, \ldots$, few nodes are randomly selected from the remaining nodes and the corresponding new links are considered as observations to the tracking algorithm. The objective of the tracking algorithm is to use the incoming observations in order to detect the hostility states of each of the remaining individuals in the network.

A detailed analogy of the experimental data related to real-world dynamic personal network is detailed as follows: The “initial nodes” which are assumed known represent an existing network, i.e., a group of people. The initial nodes are assumed known:

- It is known that whom in the group is known to whom within the group, i.e., the initial links of the network is assumed known.
- The (only) attribute of each node is assumed known, i.e., it is known if the node is hostile or not.

The simulated dynamics in the network is analogous to the following:

1) The people in the initial network make connections to other members in the network as well as new members to the network.
2) As newer members become connected to the group, the size of the network expands
3) People change their attributes influenced by their connections, i.e., a non-hostile person could turn into a hostile person and vice versa based on their connections in the network.

As the network evolves based on the above three steps, the observations about the network is made at time instances $t = 1, 2, \ldots, \inf$. The observations on a day $d$ are assumed to have the following properties:

- The observations are assumed to be collected by (let us say human) “observers.”
- Observers look for “links” in the network
- The observed links don’t have to have newly occurred on day $t$. Any existing link could have been observed on a day in the past.
- Observers are unable to view the entire links of the network. They are able to observe $r\%$ of the existing links of the network during each day.
- The observations might contain errors.

Based on the observations, the objective of the dynamic network tracking algorithm is to adaptively track the hostility
of each member of the network.

The following terminology is required in defining the performance metrics that are used in assessing the proposed approach.

- **Data variables**: actual nodes and links.
- **Inference variables**: represent attributes of nodes and links.
- **Ground truth**: True association between data variable and inference variable, e.g., \( x_{ik}^G = 1 \) if the data variable \( i \) is truly associated with the inferred variable \( k \), and \( x_{ik}^G = 0 \) otherwise.
- **Inference output**: Output from the tracking algorithm, an estimation of association between data variable and inference variable, e.g., \( x_{ik}^A = 1 \) if the tracking algorithm says that data variable \( i \) is associated with the inferred variable \( k \), and \( x_{ik}^A = 0 \) otherwise.

The following basic metrics are used in the performance evaluation.

- True positives (TP)/Hits
- True negatives (TN)/Correct rejects
- False positives (FP)/False alarms
- False negatives (FN)/Misses

where the description of each of the above basic performance metric is given respectively as

\[
TP(t) = \sum_{ik} x_{ik}^A(t) x_{ik}^G(t),
\]

\[
TN(t) = \sum_{ik} (1 - x_{ik}^A(t))(1 - x_{ik}^G(t)),
\]

\[
FP(t) = \sum_{ik} x_{ik}^A(t)(1 - x_{ik}^G(t)),
\]

\[
TN(t) = \sum_{ik} (1 - x_{ik}^A(t))x_{ik}^G(t).
\]

Finally the overall performance of the algorithm is measured in terms of accuracy defined as

\[
A(t) = \frac{TP(t) + FP(t)}{TP(t) + TN(t) + FP(t) + FN(t)}.
\]

V. CONCLUSION AND DISCUSSIONS

In this paper, we presented an approach for tracking hidden individual behaviors through longitudinal observations from the network.

We found that a hidden Markov random field (HMRF) seems to represent well the temporal as well as spatial dependency of individual attributes in the network. Further, the
adaptation of HMRF is found to provide a reasonable balance between computational complexity and modeling accuracy. We tested our tracking approach on a fictitious social network that was generated using artificial entities, groups and relationships but with a scale-free structure representing typical real-world social networks.

VI. ACKNOWLEDGEMENT

This work was funded by the Defense Advanced Research Projects Agency (DARPA). The opinions expressed here are that of the authors’ and do not necessarily reflect the policy of DARPA.
REFERENCES


