Prediction of Attributes and Links in Temporal Social Networks

Vladimir Ouzienko and **Yuhong Guo** and **Zoran Obradovic** [1]

Abstract. The analysis of social networks often assumes the time invariant scenario while in practice node attributes and links in such networks often evolve over time. In this paper, we propose a new method to predict node attributes and links in temporal networks.

1 INTRODUCTION

A typical representation model for analysis of temporal social networks consists of a series of dichotomous adjacency matrices (called sociomatrices) which define the states of a set of participating agents at each given observation time [1]. For example, to represent T temporal observations of k actors we will use T binary $k \times k$ adjacency matrices, $N^1 \cdots N^T$, where each entry $N^t_{ij} = 1$ indicates the presence of a link between the actor i and actor j at time step t; conversely $N^t_{ij} = 0$ indicates the absence of such a link. Given the social network link observations $N^1 \cdots N^T$, and the actor attribute observations $\boldsymbol{x}^1 \cdots \boldsymbol{x}^T$, where $\boldsymbol{x}^t$ is a k-length vector, we aim to make accurate predictions of the N^{T+1} and $\boldsymbol{x}^{T+1}$ in the future step.

Temporal Exponential Random Graphical Model (tERGM) [3] specifically deals with the temporal aspect of social network analysis. This model takes the Markovian assumption that each network matrix is conditionally independent of all other prior observations given its immediate prior observed matrix. The tERGM model considers only the structures and topologies of the temporal networks, while node attributes are ignored. A Hidden Temporal Exponential Random Graph Model (htERGM) [2] recovers latent temporal network structures based on attributes of observed nodes. This approach learns network structures based on node attributes, but it does not make predictions of the future step. The application of htERGM is limited to retrieval of networks of up to 10 nodes because the model requires to learn two sets of latent parameters.

In this paper, we develop a novel model that facilitates simultaneous predictions of the links and the attribute values in temporal social networks based solely on historical data. Given the evolving structure of a temporal network and the changing attribute values of the nodes, we predict the network structure and node values at the next unobserved time step. Instead of training a single joint probability prediction model, we build two conditional exponential random graph models. These two conditional predictors are mutually dependent on each other, and can then be used to predict the links, and the attribute values in an alternative way.

2 THE PROPOSED MODEL

We propose to learn directly two interdependent conditional prediction models, link and node prediction models, that can be used to predict the network structure and attribute values interdependently to avoid the expensive inference in htERGM. Our overall model will be called extended tERGM (etERGM), since the conditional models are still formulated under the similar framework as the tERGM.

2.1 Node and Link Prediction Models

For node prediction we use the following log-linear model:

$$P(\boldsymbol{x}^t|\boldsymbol{x}^{t-1}, N^t, \boldsymbol{\gamma}) = \frac{1}{Z(\boldsymbol{x}^{t-1}, N^t, \boldsymbol{\gamma})} exp\{\boldsymbol{\gamma}^{'} \boldsymbol{\psi}(\boldsymbol{x}^t, \boldsymbol{x}^{t-1}, N^t)\} Pr(\boldsymbol{x}^{t-1}) \quad (1)$$

It describes the transition of attributes from time $t-1$ to time t, conditioning on the network structure N^t at time t. Here, Z is normalization constant, and $Pr(\boldsymbol{x}^{t-1})$ is Gaussian multivariate regularization prior estimated from training data. This model encodes the dependency of the attribute values over the network structure in a direct way. The model parameter $\boldsymbol{\gamma}$ is a vector corresponding to the sufficient statistic vectors $\boldsymbol{\psi}$ which encodes the dependencies between the actors links and attributes. We used three statistics: $\psi_{sim}(\boldsymbol{x}^t, \boldsymbol{x}^{t-1}, N^t) = \sum_i^k \mathbb{I}(x_i^t, x_i^{t-1}, \sigma)$, $\psi_{dyads}(\boldsymbol{x}^t, \boldsymbol{x}^{t-1}, N^t) = k\frac{D^{'}_{count}}{D_{count}}$, $\psi_{triads}(\boldsymbol{x}^t, \boldsymbol{x}^{t-1}, N^t) = k\frac{Tr^{'}_{count}}{Tr_{count}}$. ψ_{sim} captures the temporal stability of actors' attributes. If actors attributes do not change between the observations then ψ_{sim} is large and is small otherwise. $\mathbb{I}$ is the identity function which returns 1 if node value at times t and $t-1$ is similar to σ degree, where σ is estimated from the training data. In our experiments $\mathbb{I}$ returns 1 if $|x_i^t - x_i^{t-1}| < \sigma$ where σ is a parameter, and it returns 0 otherwise. The training data was normalized to one standard deviation and we customarily set $\sigma = 0.3$. Statistics ψ_{dyads} measure the similarity of attributes for the connected nodes. It is a fraction of the total count of the linked pairs, which have similar attributes, as defined by $\mathbb{I}(x_i^t, x_j^t, \sigma)$, to the total count of directly linked pairs of the graph. Namely, D_{count} is the count of all connections in N^t and $D^{'}_{count}$ is the count of all connections in N^t subject to $\mathbb{I}(x_i^t, x_j^t, \sigma) = 1, N_{ij}^t = 1$. ψ_{triads} statistic is related to ψ_{dyads} but it operates on triads. Triads are defined as $N_{pq}^t = 1, N_{qr}^t = 1, N_{rp}^t = 1$. Tr_{count} is the count of all triads in N^t. $Tr^{'}_{count}$ is the count of triads at Tr_{count} subject to $\mathbb{I}(x_p^t, x_q^t, \sigma) = 0, \mathbb{I}(x_q^t, x_r^t, \sigma) = 0, \mathbb{I}(x_r^t, x_p^t, \sigma) = 0$.

The links are predicted by a log-linear model defined as:

$$P(N^t|N^{t-1}, \boldsymbol{x}^t, \boldsymbol{\theta}) = \frac{1}{Z(N^{t-1}, \boldsymbol{x}^t, \boldsymbol{\theta})} exp\{\boldsymbol{\theta}^{'} \boldsymbol{\psi}(N^t, N^{t-1}, \boldsymbol{x}^t, \boldsymbol{\theta}) \quad (2)$$

Similar to the tERGM model, this link prediction model defines the transition from N^{t-1} to N^t . However, different from before, we incorporate the dependency of N^t over the attributes $\boldsymbol{x}^t$ into the model

[1] Center for Information Science and Technology, Temple University, Philadelphia, PA 19122, USA. Email contact: zoran@ist.temple.edu

directly, and $\boldsymbol{\psi}(N^t, N^{t-1}, \boldsymbol{x}^t)$ denotes a list of sufficient statistics. Here, we reused the four statistics already available from tERGM. We define the following two additional statistics to capture linkage of attribute values to the network: ψ_A and $\psi_B(N^t, N^{t-1}, \boldsymbol{x}^t) = k\frac{L^{'}_{count}}{L^{''}_{count}}$. ψ_A is the same statistic as ψ_{dyads} . Statistics ψ_B is a ratio of $L^{'}_{count}$ over $L^{''}_{count}$, where $L^{''}_{count}$ is the count of links in N^t such that $\mathbb{I}(x_i^t, x_j^t, \sigma) = 1$. $L^{'}_{count}$ is the count of the links at $L^{''}_{count}$, subject to additional constraint $N_{ij}^{t-1} = 1$.

2.2 Learning Algorithm and Inference

The node and link prediction models proposed in 2.1 are both log-linear. Two sets of parameters, $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$, need to be learned there. To learn $\boldsymbol{\theta}$ we can apply Newton's optimization method [3] in straightforward fashion. Similarly to [3] we apply Newton's optimization procedure to learn parameters $\boldsymbol{\gamma}$. The main modification of our algorithm from [3] to learn $\boldsymbol{\gamma}$ is sampling procedure. In a previous study [3], Gibbs sampling was used to sample from conditional distribution of sociomatrices. Here, we replaced Gibbs sampling with the Metropolis-Hastings algorithm to sample from $P(\underline{\boldsymbol{x}}^t|\boldsymbol{x}^{t-1}, N^t, \boldsymbol{\gamma})$ distribution.

Since the node and link prediction models are interdependent, we developed the following iterative algorithm to predict the network structure and actor's attributes alternatively:

1. Initialize $\boldsymbol{x}^{t+1}$=$\boldsymbol{x}^t$
2. Do
3. Sample $\hat{N}^1 \cdots \hat{N}^B \sim P(\underline{N}^{t+1}|N^t, \boldsymbol{x}^{t+1}, \boldsymbol{\theta})$
4. $N^{t+1} = average(\hat{N}^1 \cdots \hat{N}^B)$
5. Sample $\hat{\boldsymbol{x}}^1 \cdots \hat{\boldsymbol{x}}^C \sim P(\underline{\boldsymbol{x}}^{t+1}|\boldsymbol{x}^t, N^{t+1}, \boldsymbol{\gamma})$
6. $\hat{\boldsymbol{x}}^{t+1} = mean(\hat{\boldsymbol{x}}^1 \cdots \hat{\boldsymbol{x}}^C)$
7. If $\hat{\boldsymbol{x}}^{t+1}$ and $\boldsymbol{x}^{t+1}$ have converged
8. $\boldsymbol{x}^{t+1} = \hat{\boldsymbol{x}}^{t+1}$
9. exit;
10. $\boldsymbol{x}^{t+1} = \hat{\boldsymbol{x}}^{t+1}$

The algorithm starts by initializing the attributes to the values given in the last time step, $\boldsymbol{x}^{t+1} = \boldsymbol{x}^t$. In each iteration, it first samples B sociomatrices from the link prediction model, and sets the current prediction for N^{t+1} as the average of these samples. Given the estimated N^{t+1}, the inference algorithm then samples C attribute vectors from the node prediction model and updates $\boldsymbol{x}^{t+1}$ as the mean of these vectors. This iterative process stops when it converges to static attributes estimations (usually achieved in 3-4 iterations).

3 EXPERIMENTS

To test the proposed method, we conducted experiments on two synthetic datasets and two well studied real life datasets.

The synthetic datasets are generated by defining the etERGM model that includes the node and link prediction models (1),(2) and then sampling temporal data from them. The parameters $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ used to define the etERGM model are estimated from the real-life datasets described below. We sampled a set of sociomatrices with T time points and corresponding attribute vectors; for training we kept the first $T-1$ epochs of the sampled sociomatrices and attribute vectors, and compared the predicted sociomatrix $\hat{N}^T$ and vector $\hat{\boldsymbol{x}}^T$ to underlying truth. The accuracy of structure prediction is measured as $links\_accuracy(\hat{N}^T, N^T) = 1 - \frac{\sum_{ij}^{k} |\hat{N}_{ij}^T - N_{ij}^T|}{k^2}$. The prediction of the actors' attributes is measured using the mean square error (MSE). The link prediction results were compared to a baseline predictor that takes the network structure from the immediate last step: $\hat{N}^T = N^{T-1}$. For attribute predictions, we used baseline method that assumes that the node attributes do not change between time steps : $\hat{\boldsymbol{x}}^T = \boldsymbol{x}^{T-1}$; the other alternative was to use the history mean as the prediction: $\hat{\boldsymbol{x}}^T = mean(\boldsymbol{x}^{1:T-1})$. We ran experiments on two synthetic datasets, one had 26 actors and 4 epochs, the second dataset had 50 actors and 3 epochs. The comparison results on the two synthetic datasets *Dataset1*, *Dataset2* are reported in Table 1.

We have also conducted experiments on two real life datasets *Delinquency* [5] and *Teenagers* [4]. The *Delinquency* consists of 4 temporal observation of 26 students where for each observation, the researchers collected delinquency measure (5 points scale score). The *Teenagers* consists of 3 temporal observations of 50 students, where for each observation the measurement of the students alcohol consumption was taken (also 5 points scale score). The objective was to predict the relationship network and students' attributes for each dataset. We predicted time step $t = 4$ in *Delinquency* and time step $t = 3$ in *Teenagers*. Results of our learning and inference procedures to obtain predictions also reported at Table 1. All results in Table 1 are reported averages based on 20 runs and are statistically significant (p-value < 0.05). In both experiments on synthetic and real life data the etERGM clearly outperformed the conventional predictors in prediction of actor's attributes while the difference in predicting links for *Teenagers* was inconclusive which could be expected for a network of low density where the prediction problem is very difficult.

Table 1. Evaluation on synthetic and real life datasets (*Dataset1 Dataset2 Delinquency* and *Teenagers*).

	$links\_accuracy$		Actors attributes (MSE)		
	etERGM	Last Epoch	etERGM	Last Epoch	Average
Dataset1	90.2%	88.3%	0.0188	0.0199	0.0273
Dataset2	91.3%	89.6%	0.0122	0.0182	0.0134
Delinquency	85.5%	84.9%	0.8944	1.1154	1.0128
Teenagers	83.7%	84.3%	0.7712	0.9000	0.8650

4 DISCUSSION

We have shown, that the etERGM is a viable predictor for social temporal networks links and attributes. One of its core strengths is the separate learning of its two component models, which makes it applicable to larger problems. While our approach is not directly scalable to the networks size of YouTube or Facebook, there are temporal networks that could benefit from our approach.

REFERENCES

[1] O. Frank and D. Strauss, 'Markov graphs', *Journal of the American Statistical Association*, **81**, 832–842, (1986).

[2] F. Guo, S. Hanneke, W. Fu, and E. P. Xing, 'Recovering temporally rewiring networks: a model-based approach', in *ICML '07: Proceedings of the 24th international conference on Machine learning*, pp. 321–328, New York, NY, USA, (2007). ACM.

[3] S. Hanneke and E. Xing, 'Discrete temporal models of social networks', in *Proceedings of the ICML 06 Workshop on Statistical Network Analysis*. Springer-Verlag, (2006).

[4] L. Michell and A. Amos, 'Girls, pecking order and smoking', *Social Science and Medicine*, **44**(12), 1861 – 1869, (1997).

[5] T. Snijders, C. Steglich, and G. van de Bunt, 'Introduction to stochastic actor-based models for network dynamics', *Social Networks*, (2009).