International Journal of Industrial Engineering & Production Research March 2021 Vol. 32, No. 1: 29-36
DOI: 10.22068/ijiepr.32.1.29

RESEARCH PAPER

# Developing A Method for Modeling and Monitoring of Dynamic Networks Using Latent Variables

**Fatemeh Elhambakhsh[1] & Mohammad Saidi- Mehrabad[2*]**

Received 3 May 2020; Revised 27 September 2020; Accepted 20 October 2020;


**ABSTRACT**

*Statistical monitoring of dynamic networks is a major topic of interest in complex social systems. Many researches have been conducted on modeling and monitoring dynamic social networks. This article proposes a new methodology for modeling and monitoring dynamic social networks for detection of anomalies in network structures using latent variables. The key idea behind our proposed methodology is to determine the importance of latent variables in creating edges between nodes as well as observed covariates. First, latent space model (LSM) is used to model dynamic networks. Vector of parameters in LSM are monitored through multivariate control charts in order to detect changes in different network sizes. Experiments on simulated social network monitoring demonstrate that our surveillance monitoring strategy can effectively detect abrupt changes between actors in dynamic networks using latent variables.*

**KEYWORDS:** *Latent space models; Dynamic networks; Anomaly detection; Average run length (ARL); Network surveillance.*

## 1. Introduction

Dynamic network is a type of structure that indicate communications among individuals involving relationships over time. These relationships are represented by edges between nodes. The edges between nodes change over time in dynamic networks. Therefore, the aim of dynamic social network monitoring is to identify abrupt changes in the behaviors among individuals in the network [1]. Monitoring of dynamic networks is known as network surveillance[2]. Anomalies in network structures sometimes occur because of latent and unobserved characteristics. Kim et al.[3]reviewed different types of latent variable models. The first idea of latent space models was introduced by Hoff et al.[4] that each node has positioned in a latent space. They developed a class of models where the probability of a relation between nodes depends on the positions of individuals in an unobserved space. The multiplicative form in latent space models has been studied under the name of "latent factor models"[3]. Latent factor models have some advantages when both homophily and stochastic equivalence are present in the network[5]. Hence, we used latent factor models introduced by Hoff [5] Because of their advantage. The idea of these models is based on the eigenvalue decomposition of latent variables. These papers [6,7,8,9] extended Eigen model as Hierarchical Multi linear Models. Park and Sohn [10] developed a hidden Markov multi linear tensor model (HMTM) that combines the multi linear tensor regression model with a hidden Markov model using Bayesian inference. We discuss details about Eigen model in next section. This article focuses on combining dynamic network modeling with statistical process control (SPC) techniques for detection of changes in network structures caused by latent variables. The main goal of this paper is to propose a new procedure to model and to detect changes for dynamic networks, and finally compares the performance of model using different network sizes. The overview of our proposed procedure for network modeling and monitoring is given in Fig. 1. We review some papers related to our work that used statistical process monitoring including multivariate control charts for their monitoring

* Corresponding author: *Mohammad Saidi- Mehrabad*
*mehrabad@iust.ac.ir*

1. *master degree Master student at Iran University of Science &Technology, Tehran, Iran.*
2. *PHD Prof. at College of Industrial Engineering, Iran University of Science &Technology, Tehran, Iran.*

schemes. Mazrae Farahani and Baradaran Kazemzadeh [11] proposed a statistical process control method to monitor social networks in Phase I. They used Hotelling's $T^2$ and likelihood ratio test (LRT) statistics to monitor attributes of network nodes. The probability of the number of communications were modeled through a Poisson regression model in Mazrae Farahani et al. [12]. Also, a Multivariate exponentially weighted moving average (MEWMA) and a multivariate cumulative sum (MCUSUM) control charts are used to identify significant changes over time. Heard et al. [13] used a two-stage Bayesian approach to detect anomalies in communications between individuals. Hazrati-Marangaloo and Noorossana[14] used a random graph model and two moving window-based monitoring schemes, including multipath-dependent LRT and Multivariate exponentially weighted moving average (MEWMA), to detect changes in dynamic networks. Sparks and Wilson [15] investigated detection of abrupt changes among a small team of actors in dynamic networks. They monitored teams based on the exponentially weighted moving average (EWMA) control chart. Kodali et al.[16] used dynamic latent space model (DLSM) and dynamic degree-corrected stochastic block model (DDCSBM) to generate dynamic networks. They applied summary statistics and scan method in order to detect different anomalies. Zhao et al.[17] used Priebe's scan method to monitor and analyze simulated DCSBM social network model. Yu et al.[18] proposed a multivariate network surveillance based on $\mathrm{T}^2$ control chart to monitor node propensity in dynamic DCSBM. Wilson et al.[19] illustrated the utility of the dynamic degree corrected stochastic block model (DCSBM) in simulating dynamic networks. Their proposed monitoring strategy is based on Shewhart and EWMA control charts.

The remainder of this paper is organized as follows. In the next section we briefly discuss latent space model (LSM) and its parameter estimation method using a Markov chain Monte Carlo algorithm. Then, we will describe the monitoring procedure for temporal change detection. Next, we will describe performance evaluation of proposed method using simulation study. In section 5, a numerical example will indicates the application of LSM to model real Dataset. Finally, we end with a discussion of open areas for future research in last section.

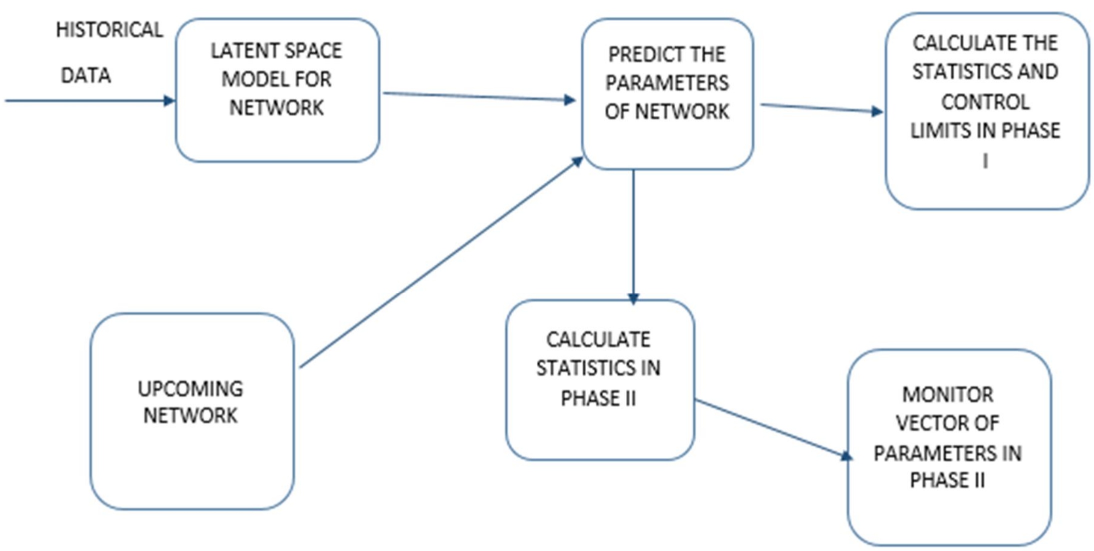



**Fig. 1. Procedure of network modeling and monitoring**

## 2. Dynamic Latent Space Model and Parameter Estimation

In this section, we discuss the details about latent Eigen model[5] as a latent variable model for modeling relational networks. Eigen model generalizes latent class models which the relationships between nodes represents by eigenvalue decomposition in latent term. Let $\{y_{i,j}: 1 \leq i \leq j \leq n\}$ denote the edge between nodes i and j, and $Y$ is defined as $n \times n$ symmetric adjacency matrix. Let $X = \{x_{i,j}, 1 \leq i \leq j \leq n\}$ indicates covariates between nodes. The model for binary network data thus becomes

$$\pi_{i,j} = \mathbf{B}^T \mathbf{x}_{i,j} + z_{i,j} \quad (1)$$

$$\Pr\left(y_{i,j} = 1 | \pi_{i,j}\right) \equiv \theta_{i,j} = \operatorname{logit}^{-1}\left(\pi_{i,j}\right) \quad (2)$$

$$\Pr\left(Y_t | \theta_{i,j}\right) = \prod_{i<j} \theta_{i,j}^{\ y_{i,j}} (1-\theta_{i,j})^{1-y_{i,j}}. \quad (3)$$

In above formulation, each node i has a vector of unobserved characteristics $\mathbf{u}_i = \{u_{i,1},\ldots,u_{i,R}\}, \mathbf{v}_j = \{v_{i,1},\ldots,v_{i,R}\}$ represent latent node positions in the $y_{i,j}$'s, and β represents coefficient of the $y_{i,j}$'s with the $x_{i,j}$'s. $\Lambda = \operatorname{diag}(\lambda_{1,t},\ldots,\lambda_{R,t})$ is the eigenvalue matrix. For more detailed information about the Eigenmodel, readers can refer to the works of [6],[7],[8],[9]. The elements of matrix z can be represented as $z_{i,j,t} = \mathbf{u}_{i,t}\Lambda\mathbf{v}_{j,t} + \varepsilon_{i,j,t}$. And, it is assumed that the error term has a standard normal distribution.

For parameter estimation, we used Hoff [5],[20] Bayesian scheme. In this regard, a Markov chain Monte Carlo (MCMC) algorithm is used to estimate the parameters. Therefore, Given a prior distribution on the model parameters, their posterior distribution are obtained via Bayes rule,

$$\Pr\left(\Pi, \mathbf{B}, \mathbf{U}, \Lambda, \mathbf{V} | Y\right) \propto \Pr\left(Y | \Pi_t, \mathbf{B}_t, \mathbf{U}_t, \Lambda_t, \mathbf{V}_t\right) \times \Pr\left(\Pi, \mathbf{B}, \mathbf{U}, \Lambda, \mathbf{V}\right) \quad (4)$$

These quantities can be approximated via Markov chain Monte Carlo sampling. In this regard, by setting starting values $\phi_0 = \{\Pi_t, \mathbf{B}_t, \mathbf{U}_t, \Lambda_t, \mathbf{V}_t\}$ a sequence of quantities $\phi_1, \phi_2, \phi_3, \ldots$ are iteratively generated as followings:

1. Sample $\mathbf{B}_t$ from its full conditional distribution,
2. For $r = 1$ to R
   - (i) Sample $u_{i,r}$, $i = 1,\ldots,n$, from its full conditional distribution,
   - (ii) Sample $v_{j,r}$, $j = 1,\ldots,n$, from its full conditional distribution,
   - (iii) Sample $\lambda_r$ from its full conditional distribution,
3. Sample $\pi^*_{i,j} = \mathbf{B}^T \mathbf{x}_{i,j} + \mathbf{u}_i \Lambda \mathbf{v}_j + \varepsilon^*_{i,j}$, where $\varepsilon^*_{i,j}$ follows normal distribution,
4. $\pi_{i,j}$ is replaced by $\pi^*_{i,j}$ with probability $\min\left(\Pr\left(y_{i,j} | \pi^*_{i,j}\right) / \Pr\left(y_{i,j} | \pi_{i,j}\right), 1\right)$.

One run of the above procedure generates a new set of parameters. The posterior mean of parameters can be approximated by the empirical mean of the MCMC samples.

## 3. Monitoring Strategy of Dynamic Networks

This section contains the details of the proposed monitoring plan. Our proposed method provides a means for estimating observed and latent parameters of dynamic networks. This section proposes a monitoring procedure for detecting sudden changes in dynamic networks caused by shifts in model parameters. In this section, suppose that we observe k sequence of networks *Y* generated under latent space model. Our proposed modeling approach estimates the parameters of dynamic network for sequence of K samples introduced in section 2. Indeed, the estimated parameters are used to predict the adjacency matrix for dynamic network. In phase I, the vector of parameters $L = \left(\beta, \bar{u}, \bar{v}, \lambda_1, \lambda_2\right)$ is fixed and estimated from an in-control sequence of networks. For each new snapshot, the vector of network parameters is predicted then the updated vector is used to predict adjacency matrix for sample k. Now we illustrate the detailed application of the proposed method. Tracy et al. [21] proposed Multivariate $T^2$ control chart based on Hotelling's T2 statistic.

$$T^2(L_i) = (L_i - \mu)'\Sigma^{-1}(L_i - \mu) \quad (5)$$

And the upper limit is defined to achieve a desired type I error ( $\alpha$ ). For phase II, new snapshots of networks are collected and their $T^2$ statistics are compared to UCL. Thus, $T^2{}_i(L) > UCL$ indicates a change in the network. Then, we use average run length (ARL) to evaluate the performance of proposed model for anomaly detection.

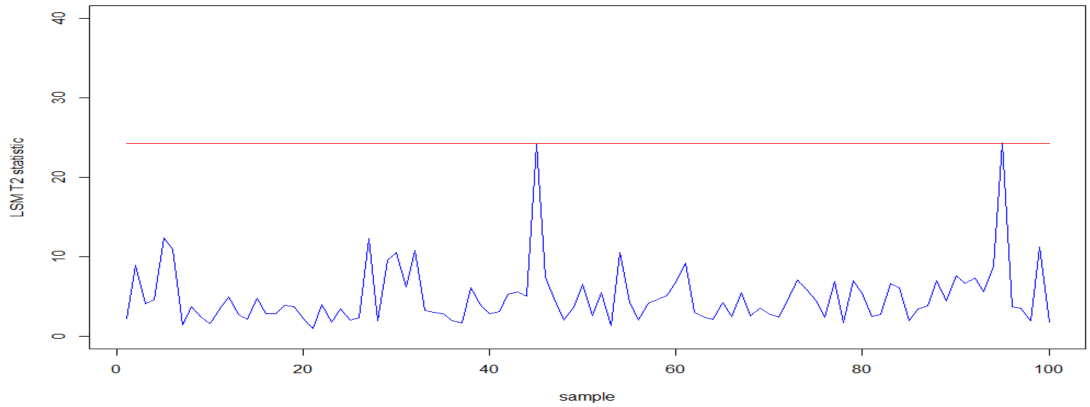
## 4. Performance Evaluation Using Simulation

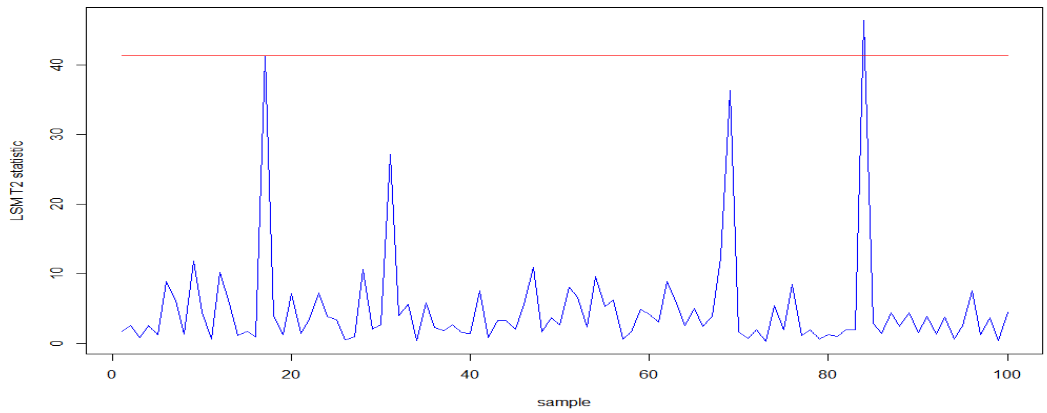
In this section, we evaluate the performance of our proposed method to compare the ability of different network sizes of LSM to detect abnormal changes by using simulation. First, we simulated 100 network samples for phase I baseline. In this setting, we consider $x_{i,j}$ as the variable between node i and j in each snapshot for LSM. Also $\beta$ is the covariate coefficient of $x_{i,j}$. Initial values and Distributions of variables and parameters are shown in table 1. L is the vector of the model parameters for each sample. Where $\beta_K$ is the intercept of $x_{i,j}$ , $u_K, v_K$ and $\lambda_{1K}, \lambda_{2K}$ are eigenvectors and eigenvalues of latent characteristics. To generate networks, we first generate a sequence of state values. Then, we calculate $\pi_{i,j}$ is calculated through the corresponding link functions. And, based on the probability in relation (2) we calculate $y_{i,j}$ .The generated networks are considered to be an input to the proposed methodology for estimating vector of parameters. Then, we calculated $T^2$ statistic for each network. Our objective is to detect abrupt changes in networks attributes caused by changes in latent variables. We repeated the whole procedure for 100 networks, and we calculated the sample mean, $\bar{L}$ ,and smaple covariance matrix. Fig. 2. shows $T^2$ control charts and control limits for phase I. For phase II, we increased the mean of normal distribution for u , and generated random network based on shifted parameter. we used average run length (ARL) for performance evaluation of networks. $ARL_0 = 100$ , and we repeated the procedure 1000 times for phase II in order to calculate $ARL_1$ for each shift. The in-control limits are shown in table 1. So, if $T^2$ statistic for each sample is greater than in-control UCL, we can say the change has occurred.

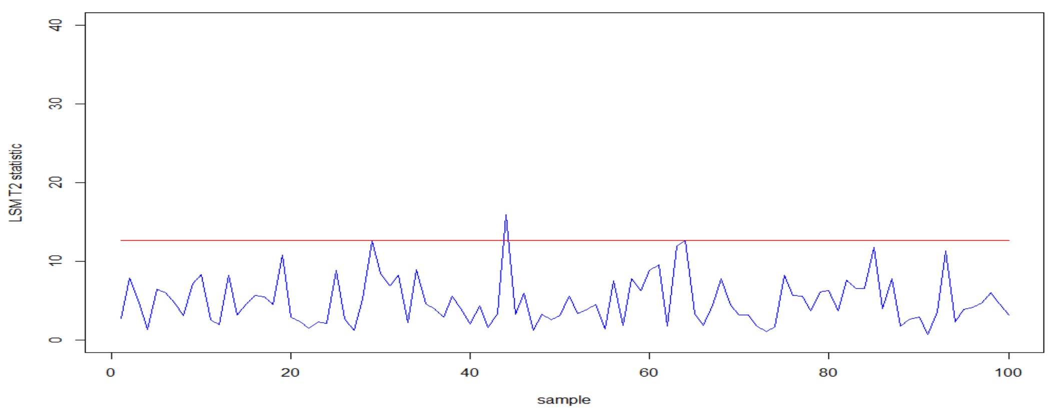
Table 2. and Fig. 3. indicate Average run length for shifts in mean of u. To compare the performance of proposed procedure in Fig 2. ARL was computed for different shifts in mean of u for different sizes. As can be seen from the Fig. 2., our proposed method can detect smaller changes in n=20 better than other ones. The networks with n=50 and 100 can detect small and bigger changes in mean U better. n=10 is not able to detect any changes well. However, latent space model with n=100 could distinguish large shifts better than others. So, we conclude that it is better to simulate large networks for modeling and monitoring networks.

**Tab. 1. Value of parameters that achieve $ARL_0 = 100$**

| n | A | UCL | R | x | u | v | $\lambda$ | $\beta_0$ |
|---|---|---|---|---|---|---|---|---|
| 10 | 0.01 | 41.29453 | 2 | N(0,1) | N(0,1) | N(0,1) | N(0,1) | 0.05 |
| 20 | 0.01 | 19.93903 | 2 | N(0,1) | N(0,1) | N(0,1) | N(0,1) | 0.05 |
| 50 | 0.01 | 12.6543 | 2 | N(0,1) | N(0,1) | N(0,1) | N(0,1) | 0.05 |
| 100 | 0.01 | 24.22434 | 2 | N(0,1) | N(0,1) | N(0,1) | N(0,1) | 0.05 |


LSM T2 statistic
sample
0
10
20
30
40
0
20
40
60
80
100


LSM T2 statistic
sample
0
10
20
30
40
0
20
40
60
80
100


LSM T2 statistic
sample
0
10
20
30
40
0
20
40
60
80
100

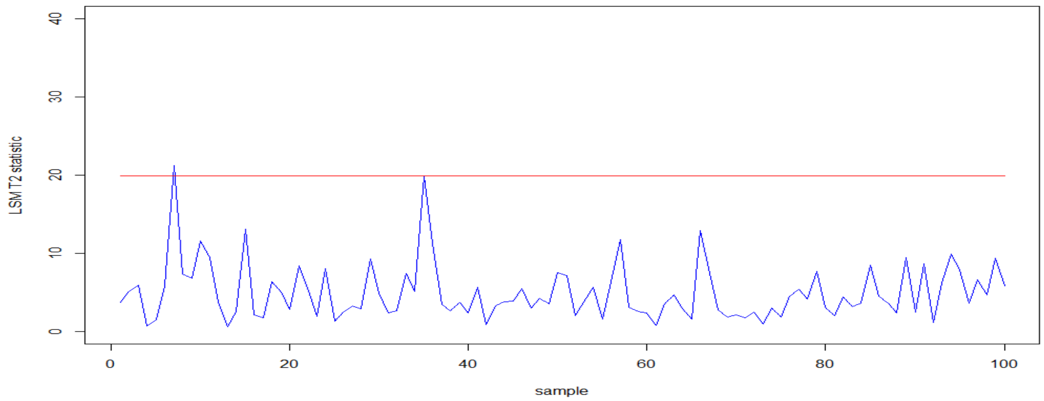



**Fig. 2. Phase I control charts for n=10, 20, 50, 100**

**Tab. 2. ARL comparison**

| Mean of u | N=10 | N=20 | N=50 | N=100 |
|---|---|---|---|---|
| 0.05 | 6.756757 | 1.142857 | 1.373626 | 1.468429 |
| 0.1 | 6.25 | 1.135074 | 1.37741 | 1.445087 |
| 0.2 | 5.847953 | 1.137656 | 1.3947 | 1.28866 |
| 0.5 | 6.289308 | 1.112347 | 1.426534 | 1.068376 |
| 1 | 6.060606 | 1.103753 | 1.371742 | 1.324503 |
| 1.5 | 4.545455 | 1.068376 | 1.371742 | 1 |
| 2 | 4.854369 | 1.077586 | 1.193317 | 1 |
| 2.5 | 4.081633 | 1.005025 | 1.162791 | 1 |

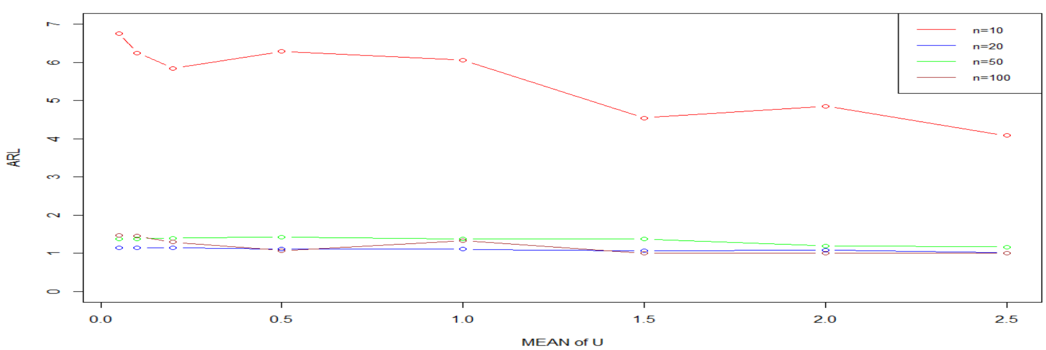



**Fig. 3. ARL performance**

## 5. Numerical Example

In this section, we illustrate the application of our proposed method on real world dataset for modeling latent space dynamic network. We used an online shopping dataset from https://www.digikala.com/opendata/. The dataset consists of 151634 customers, 95232 types of different products and 906 cities from 2014-2018 years. Fig 4. Indicates a typical monthly network for the first month of 2015. And, the amount of parameters estimation are shown in table. In this regard, an edge is placed between nodes i and j if at least node i and j bought same product during one month. And the variable X defined the same city between two nodes. As a result, the proposed method can be used to model and monitor real social network models in future research.

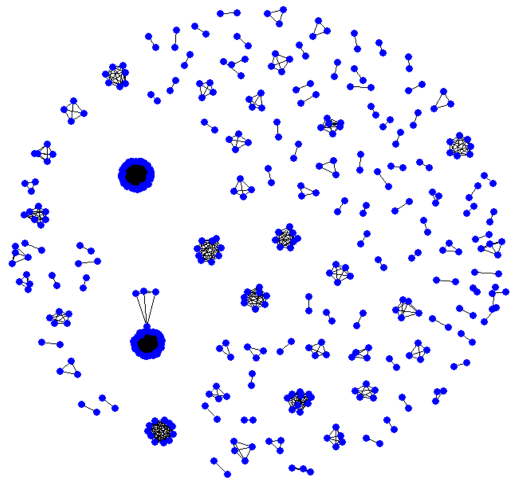


**Fig. 4. sample network of one month for 2015**

**Tab. 3. Estimated parameters for one month in 2015**

| $\beta$ | $\bar{u}$ | $\bar{v}$ | $\lambda_1$ | $\lambda_2$ |
|---|---|---|---|---|
| -0.167971037 | -0.001796568 | -0.003857942 | 0.679404339 | 0.540606557 |

## 6. Conclusion

In this paper, a new methodology for modeling and monitoring of dynamic social networks is proposed. First, we modeled dynamic networks using a LSM. Then, we estimated model parameters by Bayesian method. Next, vector of parameters monitored using multivariate control charts for detection of sudden changes in the network. In the simulation study, we generated different sizes of dynamic networks representing interactions among nodes in binary networks. The results indicated that large network sizes detect abrupt changes better than small networks. As can be seen, the networks with n=50 and 100 can detect small and bigger changes in mean of U better. However, n=10 cannot detect any changes as well as other ones. As a result, it would be better to use large networks in simulating latent networks. In future work, it would be useful to discuss and compare the use of latent space models over other models including generalized linear model (GLM)and degree corrected stochastic block model (DCSBM) for both binary and weighted networks.

## Acknowledgment

We would like to appreciate the delicate comments of the anonymous reviewers and wish to enrich the paper using these constructive comments.

## References

[1] Woodall WH, Zhao MJ, Paynabar K, et al. An overview and perspective on social network monitoring. *IISE Transactions* Vol. 49, (2017), pp. 354-365.

[2] Wilson J, Stevens N and Woodall W. Modeling and detecting change in temporal networks via a dynamic degree corrected stochastic block model, *arXiv preprint arXiv:*160504049 (2016).

[3] Kim B, Lee KH, Xue L, et al. A review of dynamic network models with latent variables. *Statistics surveys* Vol. 12, (2018), p. 105.

[4] Hoff PD, Raftery AE and Handcock MS. Latent space approaches to social network analysis. *Journal of the american Statistical association* Vol. 97, (2002), pp. 1090-1098.

[5] Hoff P. Modeling homophily and stochastic equivalence in symmetric relational data. In: *Advances in neural information processing systems* (2008), pp.657-664.

[6] Hoff PD. Hierarchical multilinear models for multiway data. *Computational Statistics & Data Analysis* Vol. 55, (2011), pp. 530-543.

[7] Hoff PD. Dyadic data analysis with amen. *arXiv preprint arXiv:150608237* (2015).

[8] Hoff PD. Multilinear tensor regression for longitudinal relational data. *The annals of applied statistics* Vol. 9, (2015), p. 1169.

[9] Hoff PD. Separable covariance arrays via the Tucker product, with applications to multivariate relational data. *Bayesian Analysis* Vol. 6, (2011), pp. 179-196.

[10] Park JH and Sohn Y. Detecting structural changes in longitudinal network data. *Bayesian Analysis* (2018).

[11] Mazrae Farahani E and Baradaran Kazemzadeh R. Phase I monitoring of social network with baseline periods using poisson regression. *Communications in Statistics-Theory and Methods* Vol. 48, (2019), pp. 311-331.

[12] Farahani EM, Baradaran Kazemzadeh R, Noorossana R, et al. A statistical approach to social network monitoring. *Communications in Statistics-Theory and Methods* Vol. 46, (2017), pp. 11272-11288.

[13] Heard NA, Weston DJ, Platanioti K, et al. Bayesian anomaly detection methods for social networks. *The Annals of Applied Statistics* Vol. 4, (2010), pp. 645-662.

[14] Hazrati-Marangaloo H and Noorossana R. Detecting outbreaks in temporally dependent networks. *Quality and Reliability Engineering International* Vol. 35, (2019), pp. 1753-1765.

[15] Sparks R and Wilson JD. Monitoring communication outbreaks among an unknown team of actors in dynamic networks. *Journal of Quality Technology* Vol. 51, (2019), pp. 353-374.

[16] Kodali L, Sengupta S, House L, et al. The Value of Summary Statistics for Anomaly Detection in Temporally-Evolving Networks: A Performance Evaluation Study. *arXiv preprint arXiv:191006497* (2019).

[17] Zhao MJ, Driscoll AR, Sengupta S, et al. The effect of temporal aggregation level in social network monitoring. *PloS one* (2018), p. 13.

[18] Yu L, Woodall WH and Tsui K-L. Detecting node propensity changes in the dynamic degree corrected stochastic block model. *Social Networks* Vol. 54, (2018), pp. 209-227.

[19] Wilson JD, Stevens NT and Woodall WH. Modeling and estimating change in temporal networks via a dynamic degree corrected stochastic block model. *arXiv preprint arXiv:160504049* (2016).

[20] Hoff PD. Multiplicative latent factor models for description and prediction of social networks. *Computational and mathematical organization theory* Vol. 15, (2009), p. 261.

[21] Tracy ND, Young JC and Mason RL. Multivariate control charts for individual observations. *Journal of quality technology* Vol. 24, (1992), pp. 88-95.



Follow This Article at The Following Site:

Elhambakhsh F, Saidi- Mehrabad M. Developing a method for modeling and monitoring of dynamic networks using latent variables. IJIEPR. 2021; 32 (1) :29-36
URL: http://ijiepr.iust.ac.ir/article-1-1064-en.html