

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188		
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 04-01-2014		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 26-Sep-2012 - 6-Oct-2013	
4. TITLE AND SUBTITLE Various New Statistical Models for Modeling and Change Detection in Multidimensional Dynamic Networks				5a. CONTRACT NUMBER W911NF-12-1-0415	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 611102	
6. AUTHORS Jing Li				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Arizona State University ORSPA PO Box 876011 Tempe, AZ 85287 -6011				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211				10. SPONSOR/MONITOR'S ACRONYM(S) ARO	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) 62271-MA-II.1	
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT In this project, we developed three new statistical models for characterizing the dynamic evolution of networks and detecting changes deviating from the normal evolving trajectory. Modeling the dynamics of evolving networks and detecting abnormal changes is of great interest in many domains. However, the fundamental methodological development in statistical modeling has been lacking. Through this project, the PI has not only significantly extended her research area and generated ample results, but also established industrial collaboration and introduced new educational materials. The PI was successfully promoted to Associate Professor with tenure, for which this					
15. SUBJECT TERMS statistical network models, dynamic networks, multi-dimension networks, change detection					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Jing Li
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			19b. TELEPHONE NUMBER 480-965-0125

Report Title

Various New Statistical Models for Modeling and Change Detection in Multidimensional Dynamic Networks

ABSTRACT

In this project, we developed three new statistical models for characterizing the dynamic evolution of networks and detecting changes deviating from the normal evolving trajectory. Modeling the dynamics of evolving networks and detecting abnormal changes is of great interest in many domains. However, the fundamental methodological development in statistical modeling has been lacking. Through this project, the PI has not only significantly extended her research area and generated ample results, but also established industrial collaboration and introduced new educational materials. The PI was successfully promoted to Associate Professor with tenure, for which this grant served as one honorable recognition.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received

Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received

Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

The work in progress is planned to be presented in 2014 INFORMS Annual Conference.

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts:

Books

Received Paper

TOTAL:

Patents Submitted

Patents Awarded

Awards

During the project performance period, the PI was promoted to Associate Professor with tenure. The PI was also given the Outstanding Junior Faculty Award by her university.

Graduate Students

<u>NAME</u>	<u>PERCENT SUPPORTED</u>	Discipline
Kun Wang	1.00	
FTE Equivalent:	1.00	
Total Number:	1	

Names of Post Doctorates

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>	National Academy Member
Jing Li	0.20	
FTE Equivalent:	0.20	
Total Number:	1	

Names of Under Graduate students supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: 0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME

PERCENT SUPPORTED

FTE Equivalent:

Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Please see attachment

Technology Transfer

ARO Final Report for STIR: Novel Network Statistics for Modern Social Interaction Modeling and Inference

Various New Statistical Models for Modeling and Change Detection in Multidimensional Dynamic Networks

Project Performance Period: 1/7~10/6/2013

PI: Jing Li, Ph.D.

Associate Professor in Industrial Engineering

School of Computing, Informatics, and Decision Systems Engineering

Arizona State University

Table of Contents

Forward	3
List of Appendices, Illustrations, and Tables	4
Statement of the Problem Studied	5
Summary of the Most Important Results	7
Appendices	10
Bibliography	16

Forward

The PI would like to sincerely thank the support on this project from the Mathematical Sciences Division of ARO and the program manager, Dr. Mou-Hsiung Chang. Significantly benefiting from the support, the PI has extended her research area to network statistics and generated ample results. The PI was successfully promoted to Associate Professor with tenure, for which this grant served as one honorable recognition. The PI also established industrial collaboration with the US Army Electronic Proving Ground to implement some of the research results to real-world applications that directly benefit the Army in defense and security. Through the successful experience of this project, the PI is enthusiastically looking forward to working with ARO on other potential projects/opportunities.

List of Appendices, Illustrations, and Tables

Appendix A: Details for the statistical formulation and solution algorithms for three proposed dynamic models

Appendix B: Details for change detection based on the three proposed dynamic models

Figure 1: Modeling the dynamics and changes in evolving networks is of great interest in many domains: (a) brain development; (b) social interaction; (c) wireless communication

Table 1: Specific aims of this project and limitations of related existing work

Table 2: Statistical formulation, solution algorithms, and properties of three proposed dynamic models for network data

Statement of the Problem Studied

In this project, we focused on developing new statistical models for characterizing the dynamic evolution of networks and detecting changes deviating from the normal evolving trajectory. Modeling the dynamics of evolving networks and detecting abnormal changes is of great interest in many domains. Figure 1 gives examples of three such domains, including (a) brain networks in developmental studies, (b) social networks in security surveillance, and (c) wireless communication networks in quality of service monitoring and intrusion detection. What complicates the problem even more is that the network measurement data is usually of mixed-type and multi-dimension. For example, brain networks may be measured by functional or structural neuroimaging techniques; social interaction may be through multi-type social media (facebook, twitter, LinkedIn, etc); wireless communication may be measured by logic link, connectivity, or connection strength. Despite the interest and importance of the problem, the fundamental methodological development in statistical modeling has been lacking. In this project, we studied the problem through three specific aims, including:

Aim 1: Multi-type multi-dimension network data fusion

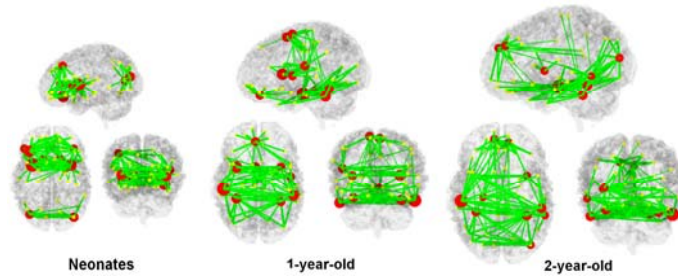
Aim 2: Dynamic modeling of evolving networks

Aim 3: Anomaly detection in evolving networks

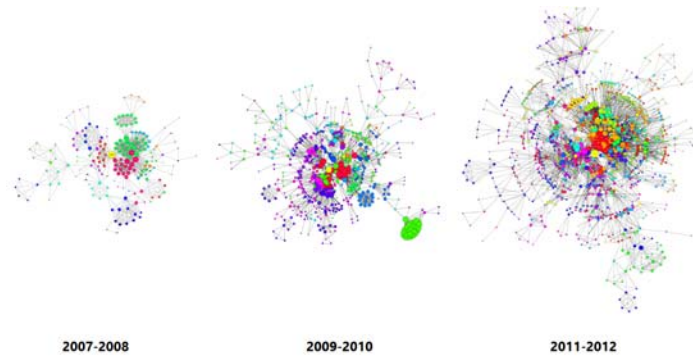
Table 1 summarizes the results of our literature review, which shows that research for addressing these specific aims is lacking and therefore this project is timely and impactful.

Table 1: Specific aims of this project and limitations of related existing work

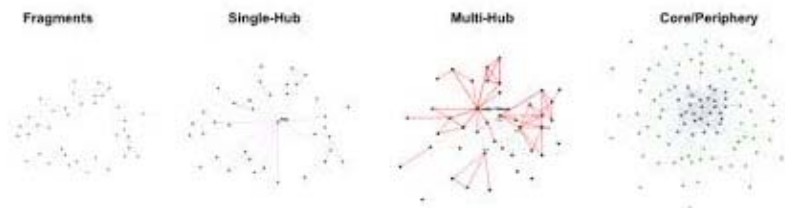
Aims	Limitations of existing work (references omitted due to space limit)
Multi-type multi-dimension network data fusion	Latent variable models focus on “variables” not “networks”. Relational learning focuses little on fusion of multi-type relations (qualitative vs. quantitative, different distributions or uncertainties).
Dynamic modeling of evolving networks	Dynamic or time-varying graphical models either assume the dynamics to be Markovian or focus on dynamic attribute data but not relational data.
Anomaly detection in evolving networks	Statistical Process Control (SPC) methods focus on monitoring of “variables” not “networks”; in network settings, SPC has only been limitedly used for network centrality measures.



(a) Evolving brain networks in child development; abnormal changes may represent developmental problems



(b) Evolving social networks; detecting abnormal changes may help social surveillance



(c) Evolving wireless communication networks; abnormal changes may indicate poor quality-of-service or intrusion

Figure 1: Modeling the dynamics and changes in evolving networks is of great interest in many domains: (a) brain development; (b) social interaction; (c) wireless communication

Summary of the Most Important Results

The important outcomes from this project are summarized in three sections: research, collaboration, and education and human resource development.

Research:

Though the original proposal of the project was written in the context of social networks, we found, during the performance period of this project, that the fundamental methodological development is needed in and also applicable to many other domains with network data such as brain networks and wireless communication networks that are closely related to the PI's past and ongoing research experience.

Aims 1 and 2: We studied extensively the characteristics of the networks in these domains and developed three new statistical models for modeling the underlying dynamics of the networks (*Aim 2*). It turned out that in two of the three models, fusion of multi-dimensional network data is just a special case of dynamic network modeling, i.e., by replacing the time index with the dimension index (*Aim 1*). Details of the model development are summarized in Appendix A. Table 2 briefly summarizes the statistical formulation and solution algorithms of the model development as well as the properties for each model.

Table 2: Statistical formulation, solution algorithms, and properties of three proposed dynamic models for network data

Proposed Models	Statistical formulation and solution algorithms	Properties
(1) A kernel smoothing estimator	The network at time t is modeled by a Gaussian Process (GP). The likelihoods of network data at different time stamps are combined by kernel weights. A Bayesian posterior estimate for the GP generates a smooth estimator for the network evolution over time. A computationally efficient approximation algorithm based on exponential-propagation (EP) was developed to solve the Bayesian posterior estimation.	The underlying assumption of this model is that the GP is smooth in time, i.e., its 1 st and 2 nd order derivatives w.r.t. time are upper-bounded by a finite constant. So this model is applicable to networks that evolve smoothly in time. It can easily incorporate attribute and multi-type, multi-dimension relational data into the modeling. The limitation is the difficulty in checking the smoothness assumption holds.
(2) A state-space framework	The networks at different time stamps are modeled by state vectors that are related to each other by a Markov dynamic model. The network data at each time stamp is related to the state vector at that time through a probability distribution that can be approximated to a Gaussian distribution by EP. An algorithm similar to optimal Bayesian filtering was developed to solve the state vector estimation. This model is	This model explicitly characterizes the network dynamics by putting a state space model on the model parameters. This provides a clean interpretation for the dynamics and also makes it easier to model different kinds of anomalies. It can easily incorporate attribute and multi-type, multi-dimension relational data into the modeling. The downside may be the computational cost of the solution algorithm since it is

	similar to the state space model in the literature but extended to a network setting.	executed recursively.
(3) A spectral evolution model	Spectral decomposition of networks at each time stamp is done with a constraint that the eigenvectors for different time stamps are related by a transformation matrix to cope with the translational, rotational and scaling effects. An optimization algorithm was developed to estimate the eigenvalues and eigenvectors at different time stamps.	This model characterizes the spectral evolution of the networks, i.e., the evolution in the eigen-space not the original space. So it provides a different but complementary perspective to the other two proposed models. This has clear interpretation that the eigenvalues capture how the density of the networks evolves and the eigenvectors capture how the structure/connectivity of the networks evolves. This model cannot easily incorporate attribute data. Efficient algorithms for spectral decomposition of large networks are yet to be developed.

Aim 3: Based on the proposed dynamic models, we further developed methods for detecting abnormal changes.

Ongoing research: We believe that we have accomplished the most challenging part – model development, during the nine-month project performance period. We are currently doing extensive simulation experiments to evaluate and compare the three proposed models. We have also found a very good resource of publicly available large network data repository, <http://snap.stanford.edu/data/index.html>. We plan to apply our models to these data and compare the results with published literature on the same data. We plan to present the research results in INFORMS 2014 Annual Conference and also submit a journal paper.

Collaboration:

In a mixer event to promote academia-defense industry collaboration, the PI presented some of the research results, which drew great attention from US Army Electronic Proving Ground (EPG). Later, the PI developed a collaborative project with EPG for modeling large mobile communication networks and monitoring and detecting poor quality of service (QoS) and potential adversarial intrusion. Mobile communication networks are commonly deployed in battlefields for soldiers to communicate with each other and with the commander for situation awareness and mission accomplishment. The QoS and security of the networks are crucial.

Education and human resource development:

Change detection or SPC for network data has not been traditionally taught in industrial engineering classes. With the abundance of network data in various application domains, there is an urgent need to introduce new methodologies for network data modeling and monitoring to students. The PI has incorporated part of the research results into the teaching materials of her advanced SPC class. In addition, the PI was successfully promoted to Associate Professor with

tenure, for which this grant served as one honorable recognition. The Ph.D. student supported by this grant also made very good progress toward his thesis.

Appendix A: Details for the Statistical Formulation and Solution Algorithms for Three Proposed Dynamic Models

We propose three different methods for modeling network dynamics.

Assume that there are N nodes in the network. Denote the nodes by $x_1, \dots, x_N$. Each node is associated with a vector of attributes. Let $\mathbf{a}_i$ denote the vector of attributes for node x_i . Let $\mathbf{E}^{(t)}$ be the observed network/relational data of the nodes at time t . $\mathbf{E}^{(t)}$ is an $N \times N$ matrix. $\mathcal{E}_{ij}^{(t)}$ denote a link between nodes x_i and x_j in $\mathbf{E}^{(t)}$. The value of $\mathcal{E}_{ij}^{(t)}$ can be 0/1 to represent the existence of a relation or a numerical number to represent the strength of a relation.

Model 1: A kernel-smoothing estimator for latent dynamic networks

Formulation: We can model the latent network dynamics by a time-varying Gaussian Process (GP). Let $\mathbf{f}^{(t)} = (f_1^{(t)}, \dots, f_N^{(t)})^T$ be random variables corresponding to nodes $x_1, \dots, x_N$ at time t . The prior distribution of $\mathbf{f}^{(t)}$ is assumed to be a zero-mean GP whose covariance matrix is defined by the attribute data, i.e., $p(\mathbf{f}^{(t)}) = GP(\mathbf{0}, \mathbf{K}_a)$ and $\mathbf{K}_a(x_i, x_j) = \exp\left(-\frac{\|\mathbf{a}_i - \mathbf{a}_j\|_2^2}{2\sigma_a^2}\right)$. The likelihood function of link $\mathcal{E}_{ij}^{(t)}$ is given as follows [1]:

$$p(\mathcal{E}_{ij}^{(t)} = 1 | f_i^{(t)}, f_j^{(t)}, \delta_i, \delta_j) = \begin{cases} 1 & \text{if } (f_i^{(t)} + \delta_i)(f_j^{(t)} + \delta_j) > 0, \\ 0 & \text{o.w} \end{cases}$$

$\delta_i \sim N(0, \sigma_\delta^2)$ denotes measurement errors. Though 0/1-type network data is considered here, we will show later that this framework allows incorporation of numerical-valued network data. Integrating out δ_i and δ_j ,

$$p(\mathcal{E}_{ij}^{(t)} = 1 | f_i^{(t)}, f_j^{(t)}) = \Phi\left(\frac{f_i^{(t)}}{\sigma_\delta}\right) \Phi\left(\frac{f_j^{(t)}}{\sigma_\delta}\right) + \left(1 - \Phi\left(\frac{f_i^{(t)}}{\sigma_\delta}\right)\right) \left(1 - \Phi\left(\frac{f_j^{(t)}}{\sigma_\delta}\right)\right). \quad (1)$$

$\Phi(\cdot)$ is the cumulative density function of the standard Gaussian distribution. Assuming the links are independent, the likelihood function of the network is $p(\mathbf{E}^{(t)} | \mathbf{f}^{(t)}) = \prod_{i,j} p(\mathcal{E}_{ij}^{(t)} | f_i^{(t)}, f_j^{(t)})$. Consider that the distribution of $\mathbf{f}^{(t)}$ changes smoothly over time. Then, the joint likelihood of the networks at all times can be written as a weighted combination of individual likelihoods and the weights are kernel functions of time, i.e.,

$$\log p(\mathbf{E}^{(1)}, \dots, \mathbf{E}^{(T)} | \mathbf{f}^{(t)}) = \sum_{s=1}^T w^{(s,t)} \log p(\mathbf{E}^{(s)} | \mathbf{f}^{(t)}),$$

where $w^{(s,t)} = \kappa\left(\frac{|s-t|}{h}\right) / \sum_{s=1}^T \kappa\left(\frac{|s-t|}{h}\right)$. According to this definition, the weight is the largest when $s = t$ and decreases as s is away from t . Using this joint likelihood function, the posterior of $\mathbf{f}^{(t)}$ is

$$p(\mathbf{f}^{(t)}|\mathbf{E}^{(1)}, \dots, \mathbf{E}^{(T)}) = p(\mathbf{f}^{(t)}) \times \prod_{s=1}^T p(\mathbf{E}^{(s)}|\mathbf{f}^{(t)})^{w^{(s,t)}}. \quad (2)$$

Estimation: The exact posterior distribution in this form is non-parametric. Though computational methods like MCMC can be used to sample from the exact posterior distribution, it is computationally too intense for large networks. We propose an approximation method based on expectation-propagation (EP) [2] to obtain an approximate Gaussian distribution for the exact posterior distribution. Specifically, we approximate $p(\mathcal{E}_{ij}^{(s)}|f_i^{(t)}, f_j^{(t)})$ by a Gaussian

distribution $r_{ij}^{(s)} \exp\left(-\frac{1}{2} \begin{bmatrix} f_i^{(t)} \\ f_j^{(t)} \end{bmatrix}^T \mathbf{\Pi}_{ij}^{(s)} \begin{bmatrix} f_i^{(t)} \\ f_j^{(t)} \end{bmatrix}\right)$ where $r_{ij}^{(s)}$ and $\mathbf{\Pi}_{ij}^{(s)}$ are successively optimized by

locally minimizing the K-L divergence. At the equilibrium the EP algorithm returns a Gaussian approximation to $p(\mathbf{E}^{(s)}|\mathbf{f}^{(t)}) \sim N(\mathbf{0}, \mathbf{\Pi}^{(s)-1})$, where $\mathbf{\Pi}^{(s)} = \sum_{i,j} \hat{\mathbf{\Pi}}_{ij}^{(s)}$. $\hat{\mathbf{\Pi}}_{ij}^{(s)}$ is $\mathbf{\Pi}_{ij}^{(s)}$ augmented to an $N \times N$ matrix. Inserting this approximate distribution to (2), the posterior distribution of $\mathbf{f}^{(t)}$ becomes $p(\mathbf{f}^{(t)}|\mathbf{E}^{(1)}, \dots, \mathbf{E}^{(T)}) \sim N(\mathbf{0}, (\mathbf{K}_a^{-1} + \sum_{s=1}^T w^{(s,t)} \mathbf{\Pi}^{(s)})^{-1})$. That is, the inverse covariance matrix of the approximate posterior Gaussian distribution is a sum of the prior inverse covariance matrix and a kernel-smoothing estimator of the inverse covariance matrix based on network data.

As a more efficient approach, $\mathbf{\Pi}^{(s)}$ can be obtained from the graph Laplacian matrix of the network. In this way, numerical-valued network data can also be incorporated using a weighted graph Laplacian matrix.

In this way, the latent dynamic networks can be characterized by the posterior inverse covariance matrix, $\mathbf{K}_a^{-1} + \sum_{s=1}^T w^{(s,t)} \mathbf{\Pi}^{(s)}$, $t = 1, \dots, T$. Note that if $T = t$, then it is an estimation problem; if $T < t$, it is a prediction problem; if $T > t$, it is an smoothing problem.

Extension: If there are multi-dimension networks at each time s , the $p(\mathbf{E}^{(s)}|\mathbf{f}^{(t)})$ in (2) can be written as $p(\mathbf{E}^{(s,1)}, \dots, \mathbf{E}^{(s,M)}|\mathbf{f}^{(t)})$, where M denotes the number of networks under consideration. $p(\mathbf{E}^{(s,1)}, \dots, \mathbf{E}^{(s,M)}|\mathbf{f}^{(t)}) = \prod_{m=1}^M p(\mathbf{E}^{(s,m)}|\mathbf{f}^{(t)})$. Then, all pervious derivations apply naturally.

Model 2: A state-space framework for latent dynamic networks

Formulation: In this formulation, the dynamics of the latent networks is modeled by a state-space model, i.e., $p(\mathbf{f}^{(t)}|\mathbf{f}^{(t-1)}, \dots, \mathbf{f}^{(1)})$. Under the simplest scenario when the Markov property holds, $p(\mathbf{f}^{(t)}|\mathbf{f}^{(t-1)}, \dots, \mathbf{f}^{(1)}) = p(\mathbf{f}^{(t)}|\mathbf{f}^{(t-1)})$. Assuming a linear Gaussian state-space model,

$$\mathbf{f}^{(t)} = \mathbf{A}^{(t-1)}\mathbf{f}^{(t-1)} + \mathbf{q}^{(t-1)},$$

$\mathbf{q}^{(t-1)} \sim N(\mathbf{0}, \mathbf{Q}^{(t-1)})$. The measurement model is the same as (1), i.e.,

$$p(\mathbf{E}^{(t)}|\mathbf{f}^{(t)}) = \prod_{i,j} p(\mathcal{E}_{ij}^{(t)} = 1|f_i^{(t)}, f_j^{(t)}). \quad (3)$$

Estimation: Following the Bayesian optimal filtering theory [3], the estimation for $\mathbf{f}^{(t)}$ can be conducted recursively as follows:

Initialization: $p(\mathbf{f}^{(0)}) \sim GP(\mathbf{0}, \mathbf{K}_a)$ defined by the attribute data

Prediction: the predicted distribution of $\mathbf{f}^{(t)}$ given data from time $s = 1, \dots, t-1$ is:

$$p(\mathbf{f}^{(t)} | \mathbf{E}^{(1:t-1)}) = \int p(\mathbf{f}^{(t)} | \mathbf{f}^{(t-1)}) p(\mathbf{f}^{(t-1)} | \mathbf{E}^{(1:t-1)}) d\mathbf{f}^{(t-1)}. \quad (4)$$

Update: given the data at time t , the posterior distribution of $\mathbf{f}^{(t)}$ can be computed by the Bayes' rule:

$$p(\mathbf{f}^{(t)} | \mathbf{E}^{(1:t)}) = \frac{1}{Z_t} p(\mathbf{f}^{(t)} | \mathbf{E}^{(1:t-1)}) \times p(\mathbf{E}^{(t)} | \mathbf{f}^{(t)}), \quad (5)$$

Z_t is a normalization constant.

We can use the EP algorithm to approximate the distribution in (3) by $p(\mathbf{E}^{(t)} | \mathbf{f}^{(t)}) \sim N(\mathbf{f}^{(t)} | \mathbf{0}, \mathbf{\Pi}^{(t)})^{-1}$. Let $p(\mathbf{f}^{(t)} | \mathbf{E}^{(1:t)}) = N(\mathbf{f}^{(t)} | \mathbf{0}, \mathbf{P}^{(t)})$. Then,

$$p(\mathbf{f}^{(t)} | \mathbf{E}^{(1:t-1)}) = N(\mathbf{f}^{(t)} | \mathbf{0}, \mathbf{A}^{(t-1)} \mathbf{P}^{(t-1)} \mathbf{A}^{(t-1)^T} + \mathbf{Q}^{(t-1)}).$$

$$\text{Furthermore, } p(\mathbf{f}^{(t)} | \mathbf{E}^{(1:t)}) \propto N(\mathbf{f}^{(t)} | \mathbf{0}, \mathbf{A}^{(t-1)} \mathbf{P}^{(t-1)} \mathbf{A}^{(t-1)^T} + \mathbf{Q}^{(t-1)}) N(\mathbf{f}^{(t)} | \mathbf{0}, \mathbf{\Pi}^{(t)})^{-1} = N(\mathbf{f}^{(t)} | \mathbf{0}, \{\{\mathbf{A}^{(t-1)} \mathbf{P}^{(t-1)} \mathbf{A}^{(t-1)^T} + \mathbf{Q}^{(t-1)}\}^{-1} + \mathbf{\Pi}^{(t)}\}^{-1}).$$

Consider a special case when $\mathbf{A}^{(t-1)} = \mathbf{I}$ and $\mathbf{Q}^{(t-1)} = \sigma_q^2 \mathbf{I}$. Then, the prediction and update in (4) and (5) become:

$$p(\mathbf{f}^{(t)} | \mathbf{E}^{(1:t-1)}) = N(\mathbf{f}^{(t)} | \mathbf{0}, \mathbf{P}^{(t-1)} + \sigma_q^2 \mathbf{I}), \text{ and}$$

$$p(\mathbf{f}^{(t)} | \mathbf{E}^{(1:t)}) = N(\mathbf{f}^{(t)} | \mathbf{0}, \{\{\mathbf{P}^{(t-1)} + \sigma_q^2 \mathbf{I}\}^{-1} + \mathbf{\Pi}^{(t)}\}^{-1}).$$

Write out a few time stamps:

$$\text{at } t = 1, p(\mathbf{f}^{(1)} | \mathbf{E}^{(1:0)}) = p(\mathbf{f}^{(0)}) \sim GP(\mathbf{0}, \mathbf{K}_a) \text{ and } p(\mathbf{f}^{(1)} | \mathbf{E}^{(1:1)}) = N(\mathbf{f}^{(1)} | \mathbf{0}, (\mathbf{K}_a^{-1} + \mathbf{\Pi}^{(1)})^{-1})$$

$$\text{at } t = 2, p(\mathbf{f}^{(2)} | \mathbf{E}^{(1:1)}) = N(\mathbf{f}^{(2)} | \mathbf{0}, (\mathbf{K}_a^{-1} + \mathbf{\Pi}^{(1)})^{-1} + \sigma_q^2 \mathbf{I}) \text{ and } p(\mathbf{f}^{(2)} | \mathbf{E}^{(1:2)}) = N(\mathbf{f}^{(2)} | \mathbf{0}, \{\{(\mathbf{K}_a^{-1} + \mathbf{\Pi}^{(1)})^{-1} + \sigma_q^2 \mathbf{I}\}^{-1} + \mathbf{\Pi}^{(2)}\}^{-1}).$$

If static networks are assumed, i.e., $\sigma_q^2 = 0$. Then, $p(\mathbf{f}^{(t)} | \mathbf{E}^{(1:t-1)}) = N(\mathbf{f}^{(t)} | \mathbf{0}, (\mathbf{K}_a^{-1} + \mathbf{\Pi}^{(1)} + \dots + \mathbf{\Pi}^{(t-1)})^{-1})$ and $p(\mathbf{f}^{(t)} | \mathbf{E}^{(1:t)}) = N(\mathbf{f}^{(t)} | \mathbf{0}, (\mathbf{K}_a^{-1} + \mathbf{\Pi}^{(1)} + \dots + \mathbf{\Pi}^{(t)})^{-1})$.

If $\mathbf{A}^{(t-1)} = \mathbf{A}$, $\mathbf{A}^T \mathbf{A} = \mathbf{I}$, and $\sigma_q^2 = 0$, then, $p(\mathbf{f}^{(t)} | \mathbf{E}^{(1:t-1)}) = N(\mathbf{f}^{(t)} | \mathbf{0}, \mathbf{A} \mathbf{P}^{(t-1)} \mathbf{A}^T)$ and $p(\mathbf{f}^{(t)} | \mathbf{E}^{(1:t)}) = N(\mathbf{f}^{(t)} | \mathbf{0}, \{\mathbf{A} \mathbf{P}^{(t-1)} \mathbf{A}^T + \mathbf{\Pi}^{(t)}\}^{-1})$.

Extension: If there are multi-dimension networks at each time t , the $p(\mathbf{E}^{(t)} | \mathbf{f}^{(t)})$ in (5) can be written as $p(\mathbf{E}^{(t,1)}, \dots, \mathbf{E}^{(t,M)} | \mathbf{f}^{(t)})$, where M denotes the number of networks under consideration. $p(\mathbf{E}^{(t,1)}, \dots, \mathbf{E}^{(t,M)} | \mathbf{f}^{(t)}) = \prod_{m=1}^M p(\mathbf{E}^{(t,m)} | \mathbf{f}^{(t)})$. Then, all pervious derivations apply naturally.

Model 3: A spectral evolution model for latent dynamic networks

Formulation: A spectral decomposition for a network is $\mathbf{E}^{(t)} = \mathbf{U}^{(t)} \mathbf{D}^{(t)} \mathbf{U}^{(t)T}$, where $\mathbf{D}^{(t)} = \text{diag}(d_1, d_2, \dots, d_K)$ is a diagonal matrix consisting of K non-zero eigenvalues of $\mathbf{E}^{(t)}$ and $\mathbf{U}^{(t)} = [\mathbf{u}_1^{(t)}, \dots, \mathbf{u}_K^{(t)}]$ consist of eigenvectors, $K \leq N$. For noisy network data, $\mathbf{D}^{(t)}$ can consist the first K significantly large eigenvalues. In network spectral analysis [4], it has been found that $\mathbf{D}^{(t)}$ characterizes the density of the network and $\mathbf{U}^{(t)}$ characterizes the structure of the network.

In conventional network spectral analysis, it assumes that there is a “constant” $\mathbf{U}$ underlying the network evolution. To extend it, our formulation allows that $\mathbf{U}^{(t)} = \mathbf{U} \mathbf{B}^{(t)}$ where $\mathbf{B}^{(t)}$ is the transformation matrix to cope with the translational, rotational and scaling effects. Given the network data $\mathbf{E}^{(1)}, \dots, \mathbf{E}^{(T)}$, modeling the network dynamics can be formulated into the following optimization problem:

$$\begin{aligned} \min_{\mathbf{U}^{(t)}, \mathbf{D}^{(t)}, \mathbf{B}^{(t)}, \mathbf{U}, t=1, \dots, T} & \left\{ \sum_{t=1}^T \left\| \mathbf{E}^{(t)} - \mathbf{U}^{(t)} \mathbf{D}^{(t)} \mathbf{U}^{(t)T} \right\|_F^2 \right\}, \\ \text{Subject to } & (\mathbf{U}^{(t)})^T \mathbf{U}^{(t)} = \mathbf{I}, \mathbf{U}^{(t)} = \mathbf{U} \mathbf{B}^{(t)} \text{ for } t = 1, 2, \dots, T. \end{aligned} \quad (6)$$

Estimation: It is not hard to derive that (6) is equivalent to the following:

$$\begin{aligned} \min_{\mathbf{U}^{(t)}, \mathbf{D}^{(t)}, \mathbf{C}^{(t)}, t=1, \dots, T} & \left\{ \sum_{t=1}^T \left\| \mathbf{E}^{(t)} - \mathbf{U}^{(t)} \mathbf{D}^{(t)} \mathbf{U}^{(t)T} \right\|_F^2 \right\} \\ \text{Subject to } & (\mathbf{U}^{(t)})^T \mathbf{U}^{(t)} = \mathbf{I}, \mathbf{U}^{(t)} = \mathbf{U}^{(t-1)} \mathbf{C}^{(t)} \text{ for } t = 1, 2, \dots, T. \end{aligned} \quad (7)$$

Because $\mathbf{U}^{(t)}$ consists of K eigenvectors orthogonal to each other, solving the optimization in (7) can be done sequentially for each eigenvector and recursively along the time, i.e.,

$$\begin{aligned} \min_{d^{(t)}, \mathbf{u}^{(t)}} & \left\| \tilde{\mathbf{E}}^{(t)} - d^{(t)} \mathbf{u}^{(t)} \mathbf{u}^{(t)T} \right\|_F^2 \\ \text{Subject to } & \mathbf{u}^{(t)T} \mathbf{u}^{(t)} = 1, \mathbf{u}^{(t)} = \mathbf{U}^{(t-1)} \mathbf{c}^{(t)} \end{aligned} \quad (8)$$

Start from $\tilde{\mathbf{E}}^{(t)} = \mathbf{E}^{(t)}$. After $\mathbf{u}^{(t)}$ and $d^{(t)}$ are obtained by solving (8), $\tilde{\mathbf{E}}^{(t)} = \mathbf{E}^{(t)} - \hat{d}^{(t)} \hat{\mathbf{u}}^{(t)} \hat{\mathbf{u}}^{(t)T}$ and (8) is solved again to obtain the next pair of eigenvalue and eigenvector. $\mathbf{U}^{(t-1)}$ is treated as constant, i.e., the estimated eigenvector matrix for the previous time stamp. Though the proof is omitted here, we have derived that the first pair of eigenvalue and

eigenvector can be obtained by the largest eigenvalue and corresponding eigenvector for matrix $\mathbf{E}^{(t)} \mathbf{U}^{(t-1)} \left(\mathbf{U}^{(t-1)T} \mathbf{U}^{(t-1)} \right)^{-1} \mathbf{U}^{(t-1)T} \left(\mathbf{E}^{(t)} \right)^T$.

Appendix B: Details for Change Detection based on the Three Proposed Dynamic Models

Change detection based on model 1: Let $\mathbf{E}^{(1)}, \dots, \mathbf{E}^{(T)}$ be the dynamic networks during the normal time period $t = 1, \dots, T$. Let t^* be a future time at which we want to determine if there is an abnormal change occurring. Based on model 1 described in Appendix A, we can derive a predicted distribution for the network at time t^* from the dynamic evolving trajectory during the normal time period, i.e.,

$$p(\mathbf{f}^{(t^*)} | \mathbf{E}^{(1)}, \dots, \mathbf{E}^{(T)}) \sim N(\mathbf{0}, (\mathbf{K}_a^{-1} + \sum_{s=1}^T w^{(s, t^*)} \mathbf{\Pi}^{(s)})^{-1}).$$

It can be further derived that the $[f_i^{(t^*)}, f_j^{(t^*)}]^T$ associated with nodes x_i and x_j follows a Gaussian distribution $[f_i^{(t^*)}, f_j^{(t^*)}]^T \sim N(\mathbf{0}, \mathbf{\Sigma}_{ij}^*)$, where $\mathbf{\Sigma}_{ij}^* = \begin{bmatrix} \kappa^*(x_i, x_i) & \kappa^*(x_i, x_j) \\ \kappa^*(x_i, x_j) & \kappa^*(x_j, x_j) \end{bmatrix}$. $\kappa^*(x_i, x_j) = \kappa(x_i, x_j) - \mathbf{\kappa}_i^T (\mathbf{I} + \mathbf{K}_a \sum_{s=1}^T w^{(s, t^*)} \mathbf{\Pi}^{(s)})^{-1} \sum_{s=1}^T w^{(s, t^*)} \mathbf{\Pi}^{(s)} \mathbf{\kappa}_j$. $\kappa(x_i, x_j)$ is the element at the i -th row and j -th column of $\mathbf{K}_a$. $\mathbf{\kappa}_i$ is a column vector corresponding to the i -th column of $\mathbf{K}_a$.

Then, the predicted distribution for a link between x_i and x_j at time t^* is:

$$p(\mathcal{E}_{ij}^{(t^*)} | \mathbf{E}^{(1)}, \dots, \mathbf{E}^{(T)}) = \int p(\mathcal{E}_{ij}^{(t^*)} | f_i^{(t^*)}, f_j^{(t^*)}) N([f_i^{(t^*)}, f_j^{(t^*)}]^T | \mathbf{0}, \mathbf{\Sigma}_{ij}^*) df_i^{(t^*)} df_j^{(t^*)},$$

which can be derived to be equal to $p(\mathcal{E}_{ij}^{(t^*)} | \mathbf{E}^{(1)}, \dots, \mathbf{E}^{(T)}) = \frac{1}{2} + \frac{\arcsin(r \mathcal{E}_{ij}^{(t^*)})}{\pi}$, where $r = \frac{\kappa^*(x_i, x_j)}{\sqrt{\kappa^*(x_i, x_i) \kappa^*(x_j, x_j)}}$. Once the network at time t^* is observed, the probability of the network

according to the normal evolving trajectory can be computed as $\prod_{ij} p(\mathcal{E}_{ij}^{(t^*)} | \mathbf{E}^{(1)}, \dots, \mathbf{E}^{(T)})$. Though more sophisticated change detection rules are to be developed, an intuitive and practically good approach is to compare this probability with that of a random network, i.e., the probability of a link existing between any two nodes is 0.5. If $\prod_{ij} p(\mathcal{E}_{ij}^{(t^*)} | \mathbf{E}^{(1)}, \dots, \mathbf{E}^{(T)})$ is bigger than a random network, then it means that there is some kind of abnormal change at time t^* .

Change detection based on model 2: Because model 2 also produces a predicted distribution for the network at a future time t^* , a similar approach to the change detection method based on model 1 can be applied. The details are omitted here because only minimum modification is needed.

Change detection based on model 3: We employ the recently developed distribution-free control chart, the Reproducing Kernel Hilbert Space (RKHS)-EWMA chart [5], for monitoring $\mathbf{U}$ and $\mathbf{D}$. Specifically, in what follows, we illustrate how the RKHS-EWMA chart can be employed

for monitoring $\mathbf{D}$, while the monitoring of $\mathbf{U}$ follows similar procedure. The estimated $\{\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(T)}\}$ and $\{\mathbf{D}^{(1)}, \mathbf{D}^{(2)}, \dots, \mathbf{D}^{(T)}\}$ from the in-control training data $\{\mathbf{E}^{(1)}, \mathbf{E}^{(2)}, \dots, \mathbf{E}^{(T)}\}$ provides an in-control library of the normal states of the process. We denote the in-control process distribution as $P_{IC}(x)$ where $\{\mathbf{D}^{(1)}, \mathbf{D}^{(2)}, \dots, \mathbf{D}^{(T)}\}$ are its samples. In the process monitoring phase, individual network data will be continuously collected, denoted as $\mathbf{E}^{(T+1)}, \mathbf{E}^{(T+2)}, \dots, \mathbf{E}^{(T+N)}$. Then, we can estimate the $\mathbf{D}^{(T+1)}, \mathbf{D}^{(T+2)}, \dots, \mathbf{D}^{(T+N)}$ and $\mathbf{U}^{(T+1)}, \mathbf{U}^{(T+2)}, \dots, \mathbf{U}^{(T+N)}$ by solving (8).

The RKHS-EWMA chart is employed to detect the process change in $\mathbf{D}^{(T+1)}, \mathbf{D}^{(T+2)}, \dots, \mathbf{D}^{(T+N)}$ by comparing them with the in-control library $\{\mathbf{D}^{(1)}, \mathbf{D}^{(2)}, \dots, \mathbf{D}^{(T)}\}$. Formerly, the process model under study can be written as

$$\mathbf{D}^{(T+n)} \sim \begin{cases} P_{IC}(x), \text{ for } n = 1, 2, \dots, \tau \\ P_{OC}(x), \text{ for } n = \tau + 1, \tau + 2, \dots \end{cases}$$

Here, τ is the time point when a process change occurs, and $P_{OC}(x)$ is the out-of-control process distribution. We let $P_t(x)$ denote the process distribution at time point t , i.e., $P_t(x)$ equals to $P_{IC}(x)$ when $t \leq \tau$, and $P_t(x)$ equals to $P_{OC}(x)$ when $t > \tau$. The essential component of process monitoring is the computation of the distance between two distributions, $D(P_{IC}, P_t)$. An out-of-control signal will be triggered if the estimated $D(P_{IC}, P_t)$ is statistically different from zero. In the estimation of $D(P_{IC}, P_t)$, many conventional SPC methods first estimate P_{IC} and P_t in either parametric or nonparametric (or distribution-free) way. However, the estimation of high-dimensional distributions has been recognized as an ill-posed inverse problem, which is generally very difficult to solve and needs a very large number of samples to guarantee an acceptable estimation uncertainty, in order to not overwhelm the subsequent estimate of the distance. In contrast to many conventional SPC methods, the RKHS-EWMA chart estimates $D(P_{IC}, P_t)$ directly, by uniquely embedding P_{IC} and P_t in the RKHS. Details of how to implement the RKHS-EWMA chart can be found in [5].

Bibliography

- [1] Chu, W., Sindhwani, V., Ghahramani, Z., and Keerthi, S.S. (2006) Relational learning with Gaussian processes. Advances in Neural Information Processing Systems (NIPS), 19, 289-296.
- [2] Minka, T.P. (2001) A family of Algorithms for Approximate Bayesian Inference. Ph.D. Thesis, MIT.
- [3] Arulampalam, M. Sanjeev, et al. (2002) A Tutorial on Particle Filters for Online Nonlinear/non-Gaussian Bayesian Tracking." Signal Processing, IEEE Transactions on 50.2, 174-188.
- [4] Cheng, Chen-Mou, H. T. Kung, and Koan-Sin Tan. (2002) Use of Spectral Analysis in Defense Against DoS Attacks. GLOBECOM'02. IEEE. Vol. 3.
- [5] Huang, S., Kong, Z.Y. and Huang, W.Z. (2013) High-Dimensional Process Monitoring and Change Point Detection Using Embedding Distributions in Reproducing Kernel Hilbert Space (RKHS). IIE Transactions, accepted (Huang, S. is a former student of the PI, recently graduated and joined the faculty of University of South Florida).