Proceedings of the 9th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'23)
Brunel University, London, United Kingdom - August 03-05, 2023
Paper No. CIST 114
DOI: 10.11159/cist23.114

Manifold Learning and Bayesian Characterization of Computer Network Traffic Supporting Machine Learning-Based Cyber System Protection

Nicholas V. Scott[1] and Jack McCarthy[2]
[1]Riverside Research Institute, Open Innovation Center, Dayton Research Center
2640 Hibiscus Way, Beavercreek, OH 45431
nscott@riversideresearch.org; jack.mccarthy@duke.edu
[2]Duke University, Department of Statistical Science
214 Old Chemistry, Durham NC 27708

Abstract

State-of-the-art cyber security rests on dynamic creation of graphical networks which model computer network traffic. This in turn allows for the assessment of optimal pathways relating different nodal components which comprise the cyber system. The estimation of such networks for cyber systems relies on robust statistical methods which provide structural understanding of cyber system components. Considering new technological initiatives directed towards using network theory as a tool to model the psychology lying behind cyber intrusion, a full top-to-bottom machine learning formulism and methodology for characterizing a computer network system is explored. The new aspect of this work is the utilization not only of classical frequentist methods for analysis but manifold learning and Bayesian methods for characterization and modelling of the interrelationships of cyber system network nodes. Such systematic analysis allows for ease of implementation of dynamic programming-based analysis directed towards optimal pathway and optimal stopping estimation.

The focus of this work is on the application of statistical methods to noise-laden, freely available 4-dimensional computer network data consisting of date, internet protocol site number, remote autonomous system numbers, and connection counts across 10 computer server sites. The objective is to demonstrate how confidence in statistical cyber behavior can be gained using multiple exploratory machine learning techniques exploited as local and global describers of cyber intrusion variability. Preliminary results show that frequentist statistics and matrix factorizations can distill Poisson probability distributions along with sub-group structure for internet protocol network traffic. Such methods can also provide insight into the relative contributions of specific internet protocol sites to global and local connection count variance. A particular sub-group of locally non-distinct, correlated internet protocol sites possess a background network structure which could be interpreted as a hidden mode of network intrusion. Manifold learning provides consistent topological results where characteristic changes in internet protocol connection counts appear near a cusp in the manifolds. Principal component analysis eigenvalue and scree plots provide evidence that only four dimensions or internet protocol sites are responsible for the global variance-based manifold. This is consistent with non-negative matrix factorization eigenmode spectral results.

Bayesian belief networks are applied as a tool to analyze the conditional probabilistic relationships existing between internet protocol server sites modeled as multi-state random variable nodes. Using the Peter and Clark algorithm for Bayesian belief network structural network learning, Bayesian belief network analysis shows specific IP sites which comprise sub-networks, including converging ones, and which parameterize the conditional probability of intrusion at server site A given server site B. The connections learned by the model are similar but not totally consistent with the correlations found in the frequentist statistics-based correlation matrix. Simulation-based instantiations at specific IP sites in the Bayesian belief network provide evidence of intrusion frequency probability, allowing information technology personnel insight into where and how resources should be placed and used to protect vital cyber systems.

Keywords: Bayesian belief networks, connection counts, cyber behavior, graphical network, Isomap, locality preserving projections, manifold learning, non-negative matrix factorization, principal component analysis, scree plot

1. Introduction

Cyber-intelligence systems need robust diagnostic and prognostic methods to promote structural understanding of computer network traffic. This allows for the creation of future innovative methods for cyber network sensing, detection, and characterization which supports safeguarding cyber systems against adversaries. Frequentist and Bayesian methods as well as manifold learning are applied to computer network traffic multidimensional data using a statistical characterization and modelling approach. Focus is on the application of these machine learning methods to noisy, four-dimensional data consisting of date, internet protocol (IP) site, remote autonomous system numbers (ASN), and connection counts in preparation for confidence assessment addressing the degree to which algorithms can detect anomalous computer network traffic behavior. In particular, manifold learning is exploited as a way to reveal latent patterns and information topology existing in the cyber data. Bayesian belief network (BBN) analysis is applied to computer network traffic to understand the statistical relationships between IP sites and to discern statistical patterns which are not immediately obvious. It is noted that the comprehension of topological cyber patterns-of-life are crucial to deducing causal adversarial mind sets responsible for attacks. In addition, the simulative power of Bayesian belief networks is responsible for the exhumation of correlative cyber structures which suggest where information technology and resources to support cyber system health and defence should be placed.

2. Data Structure and Analysis Methodologies

2.1. Data and Statistical Structure

Cyber intrusion data was downloaded from the Kaggle website [1] consisting of the aforementioned four dimensions sampled from ten IP sites over a 92-day time interval spanning the months of July 2006 to October 2006. The IP sites are labelled as IP sites 0-9. Time series of IP connection counts from IP sites 1, 3, 4, 5 and 6 are displayed in Fig. 1a showing computer site intrusion dates in late August and September 2006 at IP sites 1 and 4. These are signified by the connection count tally time series. The time series delineates the number of times per unit time each site was intruded upon by an outside source. The intrusion time points are corroborated by the local IP connection count mean value matrix shown in Fig. 1b which displays the mean connection count values over the thirteen-week time span. IP sites 1, 2, and 4 have high connection count mean values across the sample time expanse with exceptionally high values for IP sites 1 and 4 appearing during weeks 7 and 12. These high value weeks are due to the August and September intrusion times.

Local IP connection count data along with the correlation matrix displayed in Fig. 1c show strong correlation of IP sites 3, 5, and 6 suggesting coordinated cyber intrusion. Connection count histograms for each local IP site have distinctive Poisson distributions as shown in Fig. 1d. Poisson distributions describe the statistical structure of sporadic discrete computer network events occurring independently of each other with a known constant mean rate. This is a known first order model of local IP site traffic. The x-axis represents connection count intervals and the y-axis the number of connections counts from outside ASNs irrespective of which ASN the intrusion emanates from. Inspection of the shape of the histograms suggests that possible sub-groupings for the IP sites may exist. For example, qualitatively speaking histogram shapes suggests that local IP sites 3, 5, and 6 comprise one sub-group and IP sites 0, 1, 2, and 4 comprise another sub-group. Examination of the histogram shapes also suggests that IP sites 7 and 8 comprise a 3rd sub-group and IP site 9 a 4th group. Such IP site group inspection serves as a prior basis for IP site covariance or correlation analysis to be carried out using more rigorous analyses via Bayesian belief network analysis.

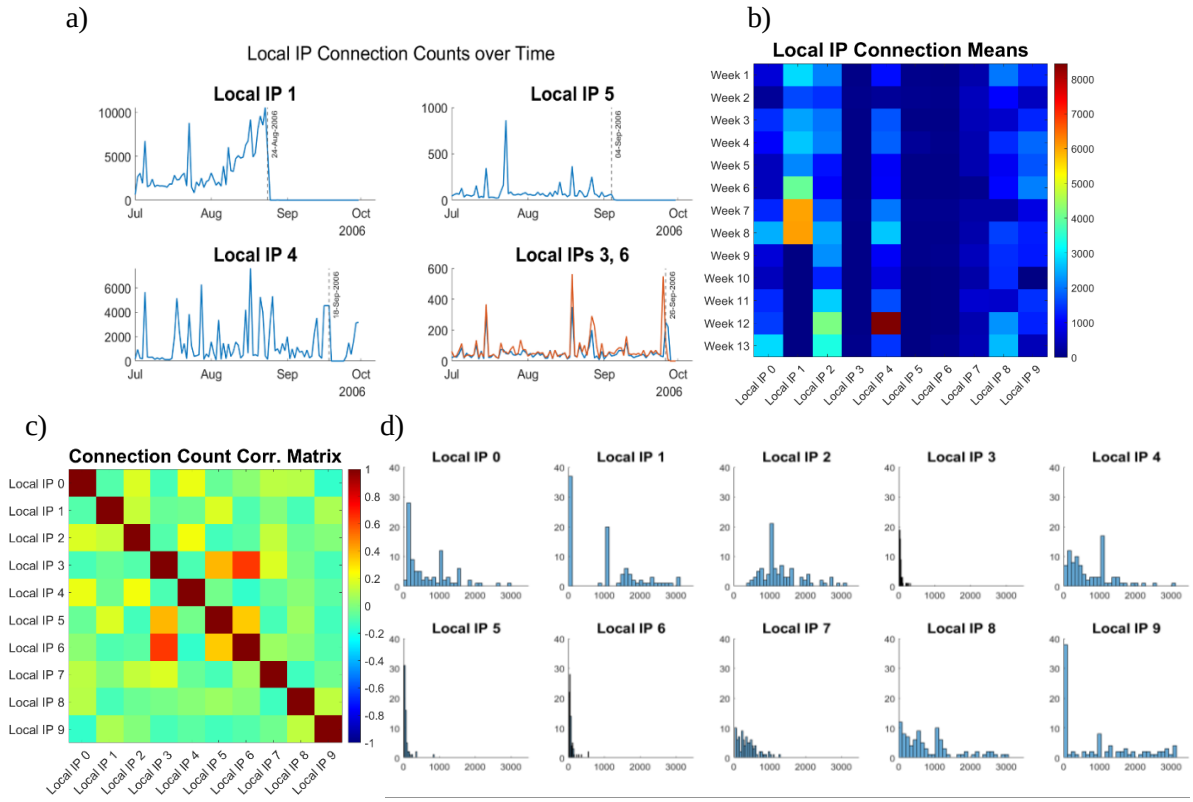

Figure 1: a) Connection count local IP traffic time series plots for IP sites 1, 3, 4, 5, and 6. The x-axis is in units of days and the y-axis is raw number of intrusions. b) Weekly mean connection counts as a function of local IP site. The x-axis and y-axis are the local IP site and the week intervals respectively. The color shaded squares are the number of cyber intrusions. c) Connection count correlation matrix for all ten local IP sites. d) Connection count histograms for all ten local IP sites. The x-axis and y-axis are the connection count label and raw number of intrusions respectively.

2.2. Principal Component Analysis, Local Preservation Projection Analysis, Nonnegative Matrix Factorization, and ISOMAP Analysis

Principal component analysis (PCA) is an eigenvalue-based matrix factorization method for decomposition of multivariate data into modes where emphasis is on preservation of global variance [2]. PCA represents data as a linear combination of basis vectors. Decomposition of multivariate data is based first on construction of the covariance matrix with the subsequent geometric transformation from an old coordinate system into a new one. In the new coordinate system, the greatest variance lies in the first projection of the data using the first eigenvector. The second greatest variance is the second coordinate or eigenvector, and so on.

Local preservation projection (LPP) is a linear dimensionality reduction method representing a linearized form of Laplacian eigenmaps [3]. The algorithm is also an eigenvalue-based decomposition where the emphasis is on local variance rather than global variance as emphasized in PCA. The LPP method is a dimensionality reduction method that is a data driven, local preservation mapping which exhumes weak covariance structure in high dimensional data. It is a form of nonlinear manifold learning which unfolds complex higher dimensional manifold structure in a lower number of dimensions for ease of visualization, preserving nearness (distance) of similar feature points. Two dimensions are often used to ease visualization where the original manifold is embedded or projected into the lower dimension while adhering to the constraint of segregating dissimilar data points [4].

A LPP algorithm mapping IP site 10-dimensional data points of the form y_i (i=1, 2, …p) can be calculated as:

$$y_i = A^T x_i \tag{1}$$

where A^T is a matrix of row LPP eigenvectors (a_0, a_1, a_2, $a_{3,\ldots}$) for each of the 10 dimensions. Data points x_i and y_i are 10-dimensional column vectors. The variable p is the number of multidimensional data points. The LPP eigenvectors, which are the rows of the matrix A^T, are called eigenfaces which allow for projection of column data vectors x_i, and which uncover low variance data structure. The sensitivity to low variance makes the transform useful for finding and segregating anomalies [5].

Nonnegative matrix factorization (NMF) is a method that factors a matrix into the product of two matrices containing nonnegative elements, and in the process, performs dimensional reduction of the information contained in the original data matrix [6]. Computer network traffic site data possessing a site dimension of l = 10 can be transformed into a data matrix X with dimension l x p. A rank r approximation to X, given by the product WH, is sought where W is a nonnegative l x r matrix and H is a nonnegative r x p matrix. The factor matrices are found by minimizing the mean square Euclidean error function:

$$f(W,H) = \frac{1}{2}\|X - WH\|^2 = \frac{1}{2}\left\|X - \sum_{j=1}^{r} \vec{w}_j \vec{h}_j^T\right\|^2 . \quad (2)$$

Here, W and H are a conglomeration of vectors delineated by $W = \left[\vec{w}_1, \vec{w}_{2,}..., \vec{w}_r\right]$ and $H = \left[\vec{h}_1, \vec{h}_{2,}..., \vec{h}_r\right]$. The NMF is accomplished by using the projected gradient local hierarchical alternating least square (HALS) algorithm [6]. The columns of W are composed of r NMF eigenvectors of dimension l that span the data space making the reconstruction of X from W and H a lower-rank approximation. The matrix H is composed of column vectors which are weights for the NMF eigenvectors.

ISOMAP is a nonlinear manifold learning technique based on the use of geodesic distance in preserving the intrinsic geometry of a multidimensional data set. It attempts to visualize the structure of multivariate data that is taken to exist in a high dimension by projection onto a low dimensional feature space. The ISOMAP algorithm also consists of three steps. The first step consists of constructing the neighborhood graph of the multidimensional data by using the k nearest neighbor approach [7]. For this problem k with the value of 8 designates the 8 data points surrounding a single data point of interest. Next, the geodesic distance between every pair of data points is calculated using Djistrka's algorithm [7]. Finally, a d-dimensional embedding of the data points is performed. The multidimensional scaling method is used to attain a low dimensional space unfolding with d=2 for easy visualization [7].

2.3. Markov Chains

Markov chains are a type of probabilistic graphical model describing changes in state of a dynamical variable over time. It is a state machine having a discrete number of states where the transitions between states are nondeterministic. The Markovian process is defined by the probability of transition P_{ij} from state q_i to another state q_j and is characterized by the Markov property where the probability of the next state only depends on the current state. They are therefore memoryless, random processes where all that is needed to predict the future is the immediate past [8]. The elements of the transition matrix can be calculated from an IP site time series by first dividing the data set into a finite number of intervals called states. The conditional probability P_{ij} is then estimated by counting the number of times the data sequence transitions into state j while being in the previous state i.

Given a Markov chain model, two questions can be posed that provide mean quantities that globally parameterize the IP connection count state machine structure. The first question is what is the percentage of time that the Markov chain is in a certain state? This question can be addressed by calculating the percentage of time statistic. This is done by first summing over the rows of the transition matrix where each row is a probability density function (pdf) for a specific state or data interval delineated by the column index N. Here N is the number of state intervals. Multiplication of the resulting N dimensional vector by 100/N yields the percentage of time pdf. The second question is what is the proportion of connection count data values in each state? The second quantity can be calculated by first taking the square of the components of the transition matrix and then summing over the rows to produce a N dimensional vector. Multiplication of the resulting N dimensional vector by 100/N yields the pdf quantifying the proportion of connection values in each state [9].

2.4. Bayesian Belief Networks Theory and Summary

Bayesian belief networks (BBNs) are probabilistic graphical models which use edges and nodes to model the joint probability distribution existing between a set of random variables describing a system [10]. They allow for statistical inferences to be made at random variable nodes when evidence is provided to one or more network nodes. Prior to statistical inference, network nodes along with nodal states need to be defined followed by structural learning which derives the directed acyclic graph (DAG) associated with the BBN. This step is concerned with exhuming the BBN topology from the data. Once the network is induced from data information, parameter learning can be performed which provides numerical values to conditional probabilities existing between nodes [11]. The defined nodes and conditional probabilities in turn allow for statistical inference where the effects of evidence at one or more random variable nodes are propagated throughout the BBN to estimate its impact on other nodes.

BBN analysis was performed on the IP connection count data using the software package Bayes Server version 9.2 manufactured by Bayes Server Ltd. which automates much of the statistical analysis including the Bayesian network structural learning and parameter learning. The Peter and Clark (PC) structural learning algorithm was implemented in this analysis which uses conditional independence to estimate a DAG associated with the IP connection count data [10]. This is a local analysis which first assumes conditional dependence between all random variable nodes which represent the ten IP sites. The algorithm subsequently breaks linkages between nodes when conditional independence is satisfied with respect to two nodes. By moving all over the IP-site domain and examining nodal linkages, the DAG is estimated. Parameter learning provides numerical values to the nodal-edge structure allowing for statistical inference between nodes. The PC algorithm uses the same relevance tree algorithm which allows for exact statistical inference rather than approximate inference [12].

3. Matrix Factorization and Manifold Learning Results

PCA and NMF analysis were performed on all ten IP sites. A three modal decomposition was performed for both analyses. The eigenmodal energy was then summed over all modes providing estimates of average PCA and NMF power spectra where the distributions delineate relative energy contributed by each IP site. Fig. 2a shows the PCA power spectrum where IP sites 1 and 4 contribute large amounts of average global variance. The NMF power spectrum, displayed in Fig. 2b, shows that IP sites 1, 2, and 4 possess large amounts of average local variance. These IP sites correspond to the same IP sites in the connection count correlation matrix that possessed exceptionally high values. IP sites 3, 5, and 6, are curiously locally non-distinct, all having low NMF power. Given that these IP sites are correlated, they could represent non-anomalous Gaussian background processes. IP attacks can be hidden in such background processes suggesting further attention to be paid at these sites by information technology (IT) personnel.

Three different manifold learning algorithms were used to reveal latent data point structure where focus is placed on the corroboration of network intrusion signatures in IP connection count data. The PCA manifold shown in Fig. 3a is the projection of the original data points using the first two PCA eigenvectors. The manifold has a sideway orientated ‘V’ structure where the top leg captures the 10-dimensional IP structure spanning weeks 1-8 and the bottom leg captures structure spanning weeks 9-13. Though manifolds are not constrained by the temporal ordering of data points, points in close spatial proximity in the manifold tend to reflect temporal proximity due to the modulation dynamics at IP sites being continuous across time. The PCA manifold experiences an abrupt change after week 8 where points from week 1-8 map to upper extremities followed by points in week 9-13 mapping to the cusp and the lower extremity of the ‘V’ shaped manifold. This pattern is conjectured to be due to the characteristic IP intrusion experienced predominantly by IP site 1 at week 9. The LPP manifold shown in Fig. 3b captures the data similarity latent structure possessing the same ‘V’ shape structure and abrupt change signature as shown in the PCA manifold. Top and bottom legs are in opposite positions with respect to the PCA manifold.

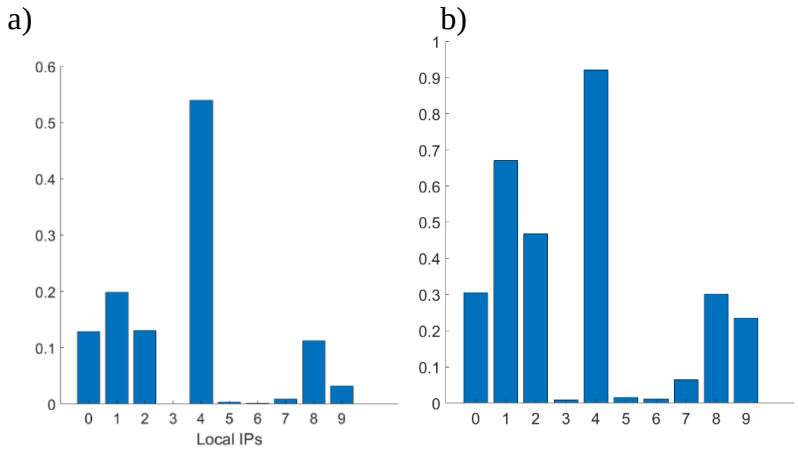

Figure 2: a) PCA average power spectrum. b) NMF average power spectrum. The x-axis delineates IP site and the y-axis the relative power averaged over 3 modes.

The ISOMAP eigenmap manifold, which utilizes geodesic distance in its estimation, has a similar 'V' shape structure where the open 'mouth' is orientated towards the left. This is shown in Fig. 4a. and corroborates the overall data topology found in the previous manifolds. Eigenvalue and scree plots shown in Fig. 4b and Fig. 4c confirm that the intrinsic dimension of the 10-dimensional data is approximately 4 through the appearance of an 'elbow' in the plots. From the results shown above, these 4 dimensions are most likely due to IP sites 0, 1, 2, and 4.

Two pdf statistics derived from the application of Markov chains can be estimated from the Markov transition matrix which parameterizes the probability of transition from one IP connection count state to another. These are the percentage or fraction of time that the Markov chain is in a certain connection count state and the proportion or fraction of connection count states that the Markov chain visits. IP sites 3, 5, and 6 possess curiously low NMF power and low connection counts and were selected for Markov chain statistical modeling analysis. Fig. 5a-c shows that IP site 3 has a uniform pdf level across connection count state intervals whereas IP sites 5 and 6 show more state tarrying at high connection count intervals. This could suggest that within this seemingly non-anomalous sub-group, there exist two IP sites which are receiving intrusions.

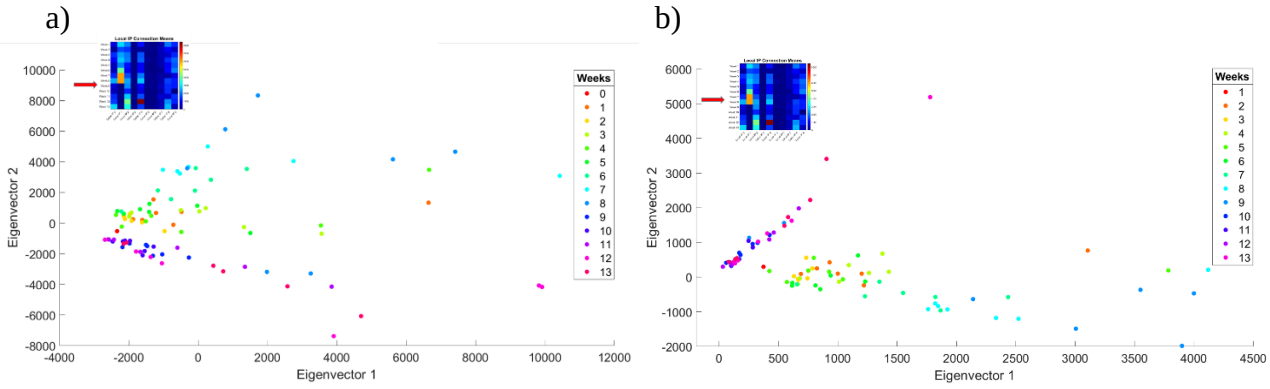

Figure 3: a) PCA eigenmap manifold. b) LPP eigenmap manifold. Eigenvector 1 and eigenvector 2 are on the x-axis and y-axis respectively. Colored data point legend designates points labeled by week. Correlation matrix shown in the upper left corner of both plots. Red arrow designates the week associated with abrupt change in the statistics captured by manifold.

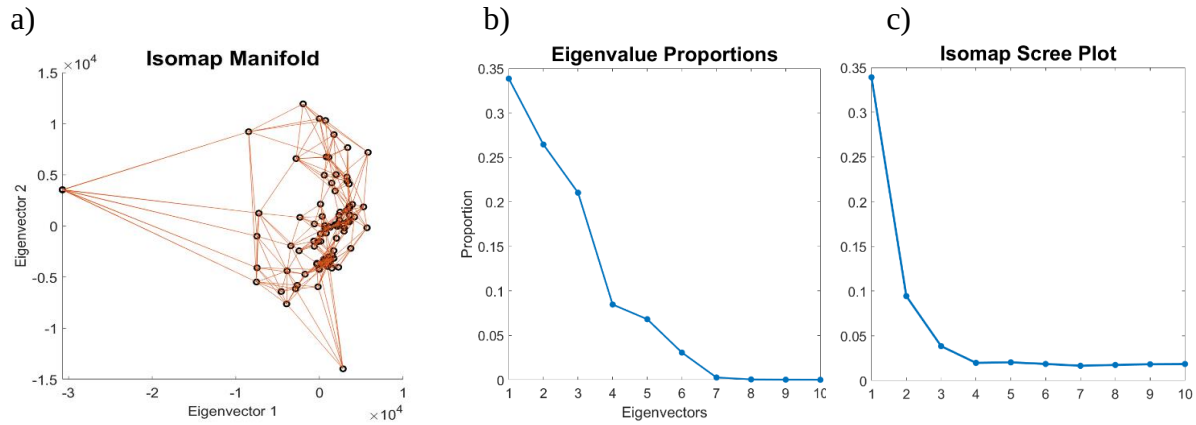

Figure 4: a) ISOMAP manifold for 10-dimensional IP connection count data. b) PCA eigenvalue plot displaying the proportion of eigenvalue energy as a function of eignvector number. c) ISOMAP scree plot identifying variance vs. intrinsic IP dimension or eigenvector.

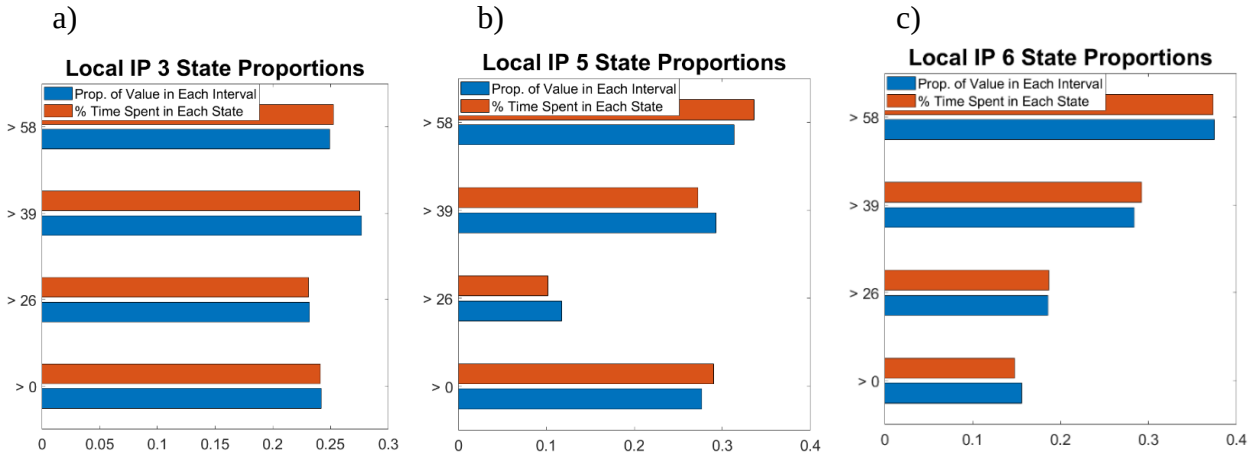

Figure 5: The proportion of IP connection count values in each state interval and the percentage of time spent in each connection count state interval for local IP sites a) 3, b) 5, and c) 6. The x-axis label is both the proportion or fraction of values in each state interval (blue) and the fraction or percentage of time in each of the four state intervals (red). The symbol, >#, represents the state interval which has a value greater than # but less than the next value in the next state interval. The highest state interval delineates values that are greater than # and less than the maximum value found by the Markov chain algorithm.

4. Bayesian Belief Network Analysis Results

The IP sites, for the purpose of this BBN analysis, are taken to be IP sites 0-9 modelled as an array of nodes where the frequency of intrusion at a site is quantified using 4 different nodal state levels. The structure of the state levels is not constant across the array but reflect local state cluster intervals for a specific node. The directed acyclic graph (DAG) represented by the BBN is a tool for understanding the modulation of nodal marginal pdfs when evidence is instantiated at one or more nodes.

Structural and parameter learning were performed, and the marginal pdf produced as shown in Fig. 6 under initial conditions containing no instantiations. The connections learned by the BBN model are similar but not totally consistent with the correlations demonstrated in the correlation matrix shown in Fig. 1c. The connection between IP sites 3 and 6 is substantiated by the BBN and the correlation matrix as well as the connections between IP sites 0 and 2, and IP sites 1 and 5. Other correlations shown in the correlation matrix do not appear as edges in the BBN model. It is worthy of note how several IP sites have no edge connections with any IP site.

Fig. 7 shows a converging sub-network where a state level 4 instantiation in IP site 2 significantly changes the marginal pdf of IP site 0 at state level 3 but causes very little change in the marginal pdf of IP site 7 at state levels of 3 and 4. The significant instantiation change displayed in IP site 0 is above 30 %. This suggests that a possible decision by IT personnel should be to check the effect of IP site 0 on IP site 2.

Fig. 8a and Fig. 8b shows how IP sites 1 and 5 significantly change each other's marginal pdfs. A state level 3 instantiation in IP site 5 is responsible for significant modulation in IP site 1 at state level 3 which is above 30%. A state level 3 instantiation in IP site 1 insignificantly modulates IP site 5 at state levels 3 and 4. This finding suggests that IP sites 1 and 5 should be examined further especially with respect to the effect of the IP site 5 child node on the IP site 1 parental node.

Fig. 9 shows a converging sub-network where high frequency intrusions (state level 4) for IP site 3 are associated with both high frequency intrusions (state level 4) in IP site 6 and low frequency intrusions (state level 1) for IP site 9. In other words, high state level intrusions for IP site 3 produced low probability state level intrusions for IP site 9 but high probability state levels for IP site 6. This suggests that IP sites 3 and 6 should be high focus areas with respect to addressing infiltrator intrusion.

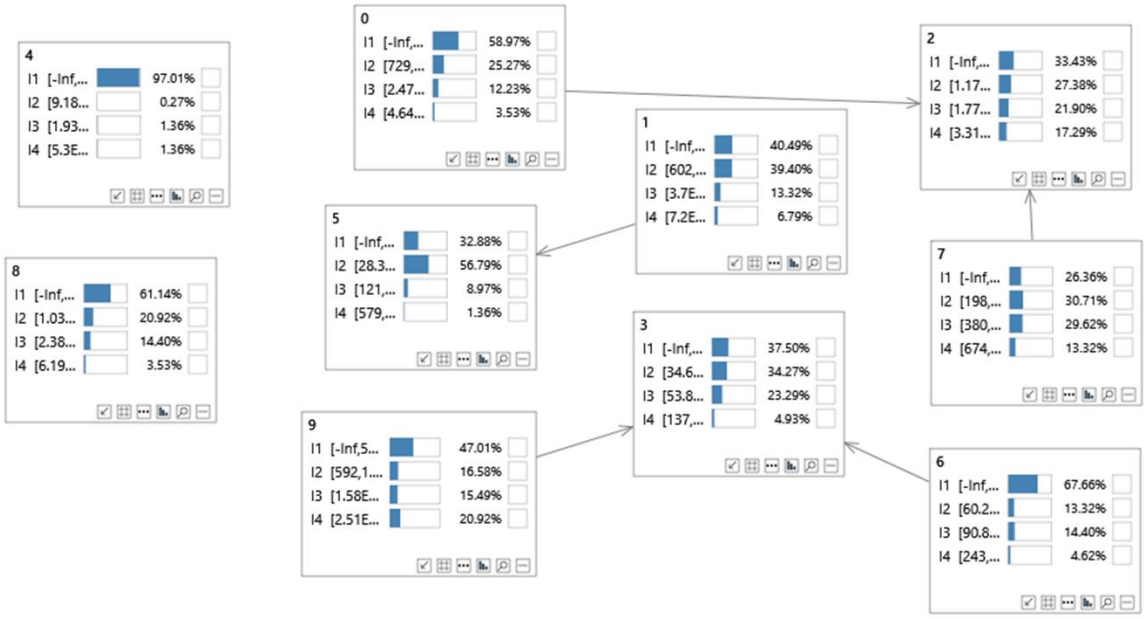

Figure 6: BBN marginal pdfs for initial conditions. IP sites numbered from 0-9. Edge connections symbolize joint or conditional pdf connections.

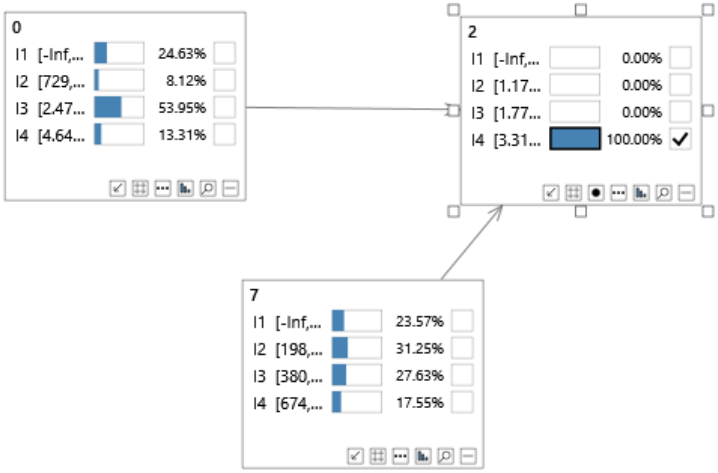

Figure 7: BBN connections for IP sites 0, 2, and 7 consist of a converging network. State level 4 instantiation for IP site 2 significantly modulates the marginal pdf of IP site 0 at state level 3 but causes an insignificant change to IP site 7 at state levels 3 and 4.

a)

b)

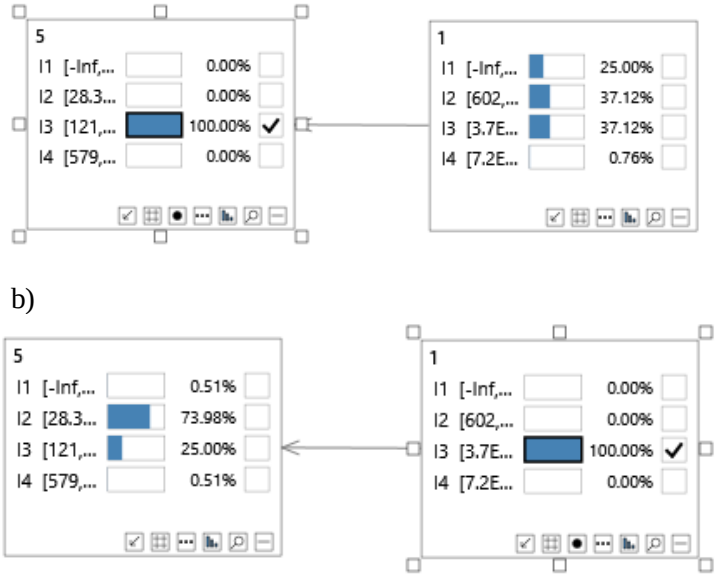

Figure 8: BBN connections for IP sites 1 and 5. a) State level 3 instantiation for IP site 5 significantly modulates the marginal pdf of IP site 1. b) State level 3 instantiation for IP site 1 insignificantly modulates the marginal pdf of IP site 5 at the medium level of state level 3.

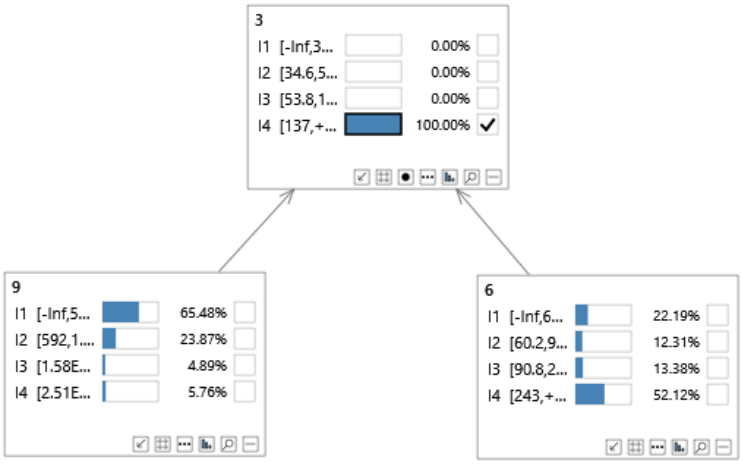

Figure 9: BBN connections for IP sites 3, 9, and 6 consisting of a converging network. State level 4 instantiation for IP site 3 produces a low intrusion frequency probabilistic state for IP site 9 at state levels 3 and 4 but produces a high intrusion frequency probabilistic state for IP site 6 at state 4 level.

Previous qualitative analysis of IP site groups performed above showed that IP site 9 was not part of any group of IP sites. This suggests that any large instantiation or high frequency of intrusion instantiation at IP site 9 should not produce any large change in IP site 3 in the BBN. Fig. 10 attempts to address the BBN relationship of IP sites 3 and 9 and shows that high frequency intrusions at IP site 9 are associated with low and mid-range frequency intrusions in IP site 3. This corroborates the idea that modulation of IP site 9 produces small levels of frequency of intrusion probability in IP site 3. From a cyber security resource perspective, these sites are probably not of great concern.

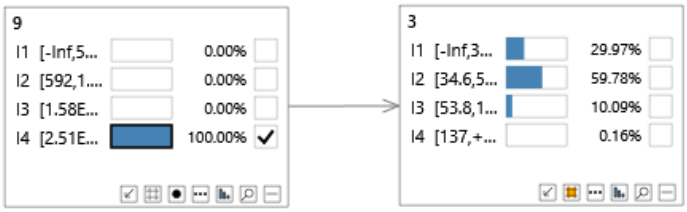

Figure 10: BBN connection for IP sites 3 and 9. IP site 9 is the parental node and IP site 3 is the child node. State level 4 instantiation for IP site 9 is associated with a high probability for medium state level 2 at IP site 3.

4. Conclusions

Matrix factorizations, manifold learning, and BBN analysis are useful analytic tools for understanding the interrelationships of computer network traffic across an array of IP sites. Matrix factorization-based power spectra can identify sites contributing significant amounts of global and local connection count variance. The identification of sites associated with low levels of local variance is also shown to suggest background processes which have the potential of hiding latent network infiltration behavior. Manifold learning can identify temporal points of significant dynamical change in the network dynamics via the embedding of multidimensional information in a lower dimensional space. Markov chain analysis allows for understand the relative tarrying in specific states for certain IP sites of interest. BBN analysis allows for lucid understanding of how connection count change at one site induces a correlative change in another. The application of BBN analysis demonstrates which sites are more highly connected to others exhuming significant sub-groups within the network. The statistical learning-based decompositions have great implications for cyber security problems where there is a desire to understand the statistical influence of high connection count activity of one IP site on another. The analyses presented here are the necessary beginning for implementation and application of more sophisticated algorithms by IT personnel which seek to understand such processes as optimal pathways, optimal cessation of movement along a pathway, and optimal selection of nodal state information over time.

References

[1] C. Crawford. (2006). Computer Network Traffic. [Online]. Available: https://www.kaggle.com/datasets/crawford/computer-network-traffic

[2] S. Marsland, *Machine Learning: An Algorithmic Perspective, Second Edition, (Chapman and Hall/CDC Machine Learning and Pattern Recognition)*. Boca Raton, FL: Chapman and Hall, 2015.

[3] A. V-Fogarassy and J. Abonyi, *Graph-Based Clustering and Data Visualization Algorithms*. New York, NY: Springer, 2013.

[4] W. Martinez and A. R. Martinez, *Computational Statistics Handbook with Matlab*. Boca Raton, FL: Chapman and Hall, 2007.

[5] J. A. Lee, and M. Verleysen, *Nonlinear Dimensional Reduction (Information Science and Statistics)*. New York, NY: Springer, 2007.

[6] A. Chichocki, R. Zdunek, A. H. Phan, and S-I. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Hoboken, NJ: John Wiley and Sons, 2009.

[7] A. J. Izenman, *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. New York, NY: Springer, 2008.

[8] P. A. Gagniuc, *Markov Chains: From Theory To Implementation and Experimentation*. Hoboken, NJ: John Wiley and Sons, 2017.

[9] O. Ibe, *Markov Processes for Stochastic Modeling*. Boston, MA: Elsevier Academic Press, 2013.

[10] K. B. Korb and A. E. Nicholson, *Bayesian Artificial Intelligence*. Florida, USA: CRC Press, 2010.

[11] U. B. Kjaerulff and A. L. Madsen, *Bayesian Network and Influence Diagrams: A Guide to Construction and Analysis*. New York, NY: Springer, 2008.

[12] A. Darwiche, *Modeling and Reasoning with Bayesian Networks*. New York, NY: Cambridge University Press, 2009.