|  |  |  |
| --- | --- | --- |
| **One day Workshop on Network and Graphical Statistics @ SOSU, ISI, 5th January 2018** | | |
| **January 5, Friday** | **Speakers** | **Chairperson** |
| 10 - 10.20 AM | Inauguration | |
| 10.20 - 10.55 AM | Soumendra Lahiri, North Carolina State University  *Statistical inference in a dynamic network model based on joint modeling of the nodes and edges* | K. S. Mallikarjun Rao (IITB) |
| 10.55 - 11.30 AM | Subhadeep Paul, Ohio State University  *Orthogonal symmetric non-negative matrix factorization under the stochastic block model* |
| 11.30 - 11.50 AM | Tea break |  |
| 11.50 AM - 12.25 PM | Snigdhanshu Chatterjee, University of Minnesota  *Networks in Climate Sciences* | Bimal K. Roy (ISI) |
| 12.25 - 1 PM | Sumanta Basu, Cornell University  *Network Modeling of High-dimensional Time Series with Applications to System-wide Risk Monitoring* |
| 1 - 2 PM | Lunch |  |
| 2 - 2.35 PM | Shyam Ranganathan, Virginia Tech University  *Community detection in multiplex networks* | Murari Mitra (IIEST) |
| 2.35 - 3.10 PM | Sandipan Roy, University College London  *Nonparametric Change-point Detection in Dynamic Networks* |
| 3.10 - 3.30PM | Tea break |  |
| 3.30 - 4.05 PM | Sabyasachi Chatterjee, University of Illinois at Urbana Champaign  *Estimation in Tournaments and Graphs with Monotonicity Constraints* | Debashis Paul (UC Davis) |
| 4.05 - 4.40 PM | Srijan Sengupta, Virginia Tech University  *SONNET: Efficient community detection via network subsampling* |
| 4.40 – 5PM | Valedictory | |

Venue: SOSU Seminar Room, C D Desmukh Bhavan 2nd floor

**Soumendra Nath Lahiri**

Department of Statistics, North Carolina State University

Title: Statistical inference in a dynamic network model based on joint modelling of the nodes and edges

Abstract: We consider a dynamic network where at any given point of time, the probability of formation of an edge is determined by the past history as well as by node properties. We consider estimation of the model parameters and their large sample properties.

**Subhadeep Paul,**

Department of Statistics, The Ohio State University

Title:  Orthogonal symmetric non-negative matrix factorization under the stochastic block model

Abstract: We present a method based on the orthogonal symmetric non-negative matrix tri-factorization of the normalized Laplacian matrix for community detection in complex networks. While the exact factorization of a given order may not exist and is NP hard to compute, we obtain an approximate factorization by solving an optimization problem. We establish the connection of the factors obtained through the factorization to a non-negative basis of an invariant subspace of the estimated matrix, drawing parallel with the spectral clustering. Using such factorization for clustering in networks is motivated by analyzing a block-diagonal Laplacian matrix with the blocks representing the connected components of a graph. The method is shown to be consistent for community detection in graphs generated from the stochastic block model and the degree corrected stochastic block model. Simulation results and real data analysis show the effectiveness of these methods under a wide variety of situations, including sparse and highly heterogeneous graphs where the usual spectral clustering is known to fail. Our method also performs better than the state of the art in popular benchmark network datasets, e.g., the political web blogs and the karate club data.

**Snigdhansu Chatterjee**

School of Statistics, University of Minnesota

Title: Networks in Climate Sciences

Abstract: The variables of interest in climatological studies typically exhibit two kinds of directed patterns: one that characterizes local or near-neighbor dependencies and another that characterizes long-distance relations or tele-connections. We study different ways of eliciting such climate network patterns, and study some of their statistical properties. We consider the spatial relationship between atmospheric pressure at sea level as a case study.

**Sumanta Basu**

Department of Biological Statistics and Computational Biology & Department of Statistical Science, Cornell University, <http://faculty.bscb.cornell.edu/~basu/>

Title: Network Modeling of High-dimensional Time Series with Applications to System-wide Risk Monitoring

Abstract: Measuring connectedness among financial institutions is central in many aspects of financial economics, including system-wide risk monitoring and identifying systemically risky institutions. In this work, we present a unified framework for measuring connectivity among firms or asset classes from multivariate time series data. The proposed framework relies on regularized estimation of high-dimensional vector autoregressive models (VAR), is flexible enough to incorporate grouping and latent structures among firms, allows parallel implementation for large data sets and enjoys strong statistical guarantees under high-dimensional scaling. We apply our method to analyze connectivity among stock returns of leading financial firms in the U.S. before, during and after the financial crisis of 2007-2008, and demonstrate promising results in detecting important systemic events and systemically risky institutions.

**Shyam Ranganathan**

Virginia Tech University

Title: Community detection in multiplex networks

Abstract: Community detection in networks is a well-studied problem and a number of solutions have been proposed in the literature. In addition, a lot of work has also gone into understanding the problem from an inferential perspective, especially using Stochastic Block Modelling (SBM). Recently, researchers have also started looking at the community detection problem for multiplex networks. A simplistic notion of multiplex networks or multigraphs occurs as a natural extension of single-layer networks, with different edges between nodes now being defined on different layers. In this talk, we will introduce a multiplex network that is of interest to economists - the trade and financial networks linking different countries. We have data on trade and financial linkages between a number of countries over a period of time (leading actually to a weighted, directed, dynamic multiplex network!) and the application we are interested in is to find how any disturbance in the financial network (e.g.. the Great Recession) could affect the trade network (e.g., slowdown in global trade) or vice versa. The first step in this direction is to understand the empirics of these networks, and we use community detection on this multiplex network, initially separate algorithms on the two layers, to uncover patterns such as trade blocs (e.g., NAFTA), and financial blocs (e.g., the Scandinavian "banking union” within Europe). This helps understand the contagion on the networks as a function of community memberships. But, an essential aspect of community detection on the multiplex network should account for the effects of both layers of the multiplex network simultaneously in order to model real-world econo-politics. To this end, we explore multilayer SBM (MSBM) and latent space approaches to perform multiplex network community detection and we will present preliminary results at the workshop.

**Sandipan Roy**

University College London

Title: Nonparametric Change-point Detection in Dynamic Networks

Abstract: This paper introduces a new model for networks evolving over time. We model the link probabilities among network nodes using a nonparametric spatio-temporal function that extends the idea of graphons for static networks. We consider a change-point detection problem for a particular class of this non-parametric time-varying model viz. a non-stationary time evolving network model that has a structural break at a certain time-point. In order to detect the change-point we employ a testing procedure based on a discrepancy measure between time varying graphon functions pre and post change-point. We study the asymptotic properties of our test-statistic and find an approximation to the type-I error of change detection. The performance of our methodology is carried out on synthetic datasets and on two human interaction datasets collected over time.

**Sabyasachi Chatterjee**

University of Illinois at Urbana Champaign

Title: Estimation in Tournaments and Graphs with Monotonicity Constraints

Abstract: We consider the problem of estimating the probability matrix governing a tournament or linkage in graphs from incomplete observations, under the assumption that the probability matrix satisfies natural monotonicity constraints after being permuted in both rows and columns by some latent permutation. We propose a natural estimator which bypasses the need to search over all possible latent permutations and hence is computationally tractable. We then derive asymptotic risk bounds for our estimator. Pertinently, we demonstrate an automatic adaptation property of our estimator for several sub classes of our parameter space which are of natural interest, including generalizations of the popular Bradley Terry Model in the Tournament case, the β model and Stochastic Block Model in the Graph case, and Hölder continuous matrices in the tournament and graph settings.

**Srijan Sengupta**

Virginia Tech University

Title: SONNET: Efficient community detection via network subsampling

Abstract: Statistical network analysis is a fast growing research area with diverse applications spanning several scientific disciplines. The community structure observed in networks has been of particular interest in the statistics literature, along with the closely related task of community detection, i.e., discovering similar groups of nodes. Popular community detection methods include clustering-based methods like Spectral clustering, and modularity maximization methods like Extreme points. Such methods have excellent statistical properties but are often computationally expensive making them prohibitively costly particularly for large networks. We propose SONNET, a general subsampling strategy based on overlapping sub-networks, that is computationally scalable and can substantially speed up a remarkable range of community detection methods with little loss of statistical accuracy. We demonstrate the advantages of our method using spectral clustering, heterogeneous spectral clustering, and extreme points, on both simulated networks and benchmark network datasets like political blogs and DBLP.