BMS Summer School 2017

Probabilistic and statistical methods for networks

TU Berlin, Math Building, Straße des 17. Juni 136
21 August - 1 September, 2017

(as at 11 August, 2017)

Lecture Programme:

- **Shankar Bhamidi** (North Carolina):
  Probabilistic and statistical problems pertaining to dynamic networks
  Exercise by: Suman Chakraborty
- **Benedikt Jahnel** (Berlin):
  Stochastic geometry in telecommunications
  Exercise by: András Tóbiás
- **Max Klimm** (Berlin):
  Selfish routing in networks
  Exercise by: Philipp Warode
- **Peter Mörters** (Bath/Cologne):
  Reinforced branching processes
  Exercise by: Anna Senkevich
- **Tiago Peixoto** (Bath):
  Statistical inference of network structure and dynamics
- **Jörg Polzehl** (Berlin):
  Connectivity networks in neuroscience – construction and analysis
- **Wilhelm Stannat** (Berlin):
  Stochastic mean-field theories for brain networks
- **Lenka Zdeborová** (Saclay):
  Inference on networks via cavity method and message passing

Organisation:

- Registration is in front of MA041.
- All lectures, exercises and talks take place in MA041.
- All coffee breaks take place in MA144.
- The Pizza Welcome on Monday (21 Aug) 18:00 takes place in the BMS lounge, MA212.
- The exercises will last for 90 minutes and will have a ten-minutes break inbetween.
- The research talks by organisers, invitees and participants will have 45 minutes including discussion.

Scientific Organisers:

- Wolfgang König (Berlin)
- Cécile Mailler (Bath)
- Marcel Ortgiese (Bath)
- Matt Roberts (Bath)
- Tim Rogers (Bath)

BMS One-Stop Office:

`summerschool@math-berlin.de`
Annika Preuß: MA219
+49 - 30 - 314-78611

The lectures and their abstracts:

Shankar Bhamidi (North Carolina):

Probabilistic and statistical problems pertaining to dynamic networks

The last few years have witnessed an explosion in the amount of empirical data on real networks motivating an array of mathematical models for the evolution of such networks. Examples range from biological networks (brain networks of interacting neurons), information transmission (Internet), transportation, social networks and swarm intelligence and the evolution of self-organized behavior through the interactions of simple agents. This has stimulated vigorous activity in a multitude of fields, including biology, statistical physics, statistics, mathematics and computer science to understand these models and quantify their predictions and relevance to real systems.

The aim of this course is to introduce junior researchers to one corner of this vast field, Dynamic networks: systems that evolve over time through probabilistic rules. The following two main themes will be pursued:

- Emergence of macroscopic connectivity [2 lectures]: The first two lectures will delve into techniques for understanding how macroscopic connectivity in the network arises via microscopic interactions between agents in the network. We will consider various random graph models and study the nature of emergence of the giant component, in particular establishing sufficient conditions for these objects to belong to the same universality class as governed by Aldous's multiplicative coalescent. We will show how these techniques can be used to study not just sizes of maximal components in the critical regime but also show that the maximal components appropriately scaled converge to limiting random metric spaces.
- Evolving networks and continuous time branching [1 lecture]: The last lecture will study the so-called preferential attachment family of network models emphasizing one particular technical tool: continuous time branching processes. We will show how this technique leads to rigorous asymptotic descriptions to a number of problems in the statistical modeling of real world systems ranging from Twitter event networks to change point detection in evolving networks.

Benedikt Jahnel (Berlin):

Stochastic geometry in telecommunications

We will introduce basic concepts in continuum stochastic geometry such as percolation or chemical distance. We will then use these concepts to derive connectivity properties of ad-hoc telecommunication networks with multiple structural components.

Suggested reading materials: R. Meester and R. Roy. Continuum Percolation. Cambridge University Press, Cambridge, 1996.

Max Klimm (Berlin):

Selfish routing in networks

Many networks such as road and telecommunication networks are used by a multitude of users that follow their own interests. In these networks it is impossible to globally impose routing strategies that are optimal for the overall network performance. Instead, the state of the system is largely determined by the individual routing strategies of the network users. Such selfish behavior poses a number of important questions that are addressed in this course. First, we will discuss the issue of network stability. Specifically, we will be interested in the questions whether there exist equilibrium points in the network where all users are satisfied with their current routing decisions, and how the users can reach these equilibria. Second, we will examine the degradation of the network performance due to the users' uncorrdinated behavior by comparing the network performance in a globally optimal state with the network performance in an equilibrium.

Peter Mörters (Bath/Cologne):

Reinforced branching processes

In this minicourse we look at a class of stochastic processes with reinforcement, which can be described in terms of general branching processes, also known as Crump-Mode-Jagers processes. In the first part of the course I will explain the classical convergence theory of these processes under the assumption of existence of a Malthusian parameter. In the second part I will present recent research (joint with Steffen Dereich and Cécile Mailler) exploring the interesting phenomena that may occur if this assumption fails. The theory has applications to urn schemes, networks, and genetic house-of-cards models, some of which will be discussed in the course.

Tiago Peixoto (Bath):

Statistical inference of network structure and dynamics

Networks are shaped by evolutionary mechanisms, and determine the central aspects of how complex systems function. However, differently from systems that are naturally embedded in space, we cannot simply 'look' at network in order to extract its most important structural patterns. Instead, we must rely on well-founded algorithmic methods to extract this information from data in an interpretable way. In this lecture, we review a principled approach to this problem based on the elaboration of probabilistic models of network structure, and their statistical inference from empirical data. We aim to cover the following topics:

- The stochastic block model (SBM) and its variants (degree correction, overlapping groups, etc.)
- Bayesian inference and model selection: Distinguishing structure from noise.
- Generalizing from data: Prediction of missing and spurious links.

- Model extensions: Layered, dynamic SBMs, and generalized models on continuous latent spaces.
- Fundamental limits of inference: The undetectability transition
- Efficient inference algorithms.

Jörg Polzehl (Berlin):

Connectivity networks in neuroscience – construction and analysis

In the lectures we will introduce the basic concepts of anatomical, effective and functional connectivity in neuroscience. We will describe the neuro-imaging experiments providing the imaging data and models used in the statistical analysis of these data, construction of connectivity networks and analysis of their properties. Hence, the two lectures will evolve around:

- Principles of fMRI and connectivity networks in neuroscience
- Statistical models for functional connectivity networks

Wilhelm Stannat (Berlin):

Stochastic mean-field theories for brain networks

Biological neural systems provide an important field of application for the mathematical theory of networks. In this case nodes stand for the individual neuron and the edges for directed interneuron connections. Neural activity can then be modelled in terms of the membrane potential attached to each neuron and neural processing in its simplest form can be realized as a point event, called a spike, emitted from one neuron through the network.

The crucial feature of all neural activity is the fact that the emission of spikes of a single neuron is a highly nonlinear and variable functional of its received input and that only little is known on the statistical properties of the network architecture, so that the impact on the resulting population dynamics is only poorly understood.

In my lectures I will first introduce the basic mathematical building blocks of neural networks modelling neural activity in biological neural networks, ranging from purely statistical models, in particular binary neural models and stochastic integrate-and-fire models, up to detailed biophysical models, and shortly discuss their relevant mathematical properties with regards to population activity in networks. In the next step I will introduce the mathematical concepts to formulate dynamical equations for macroscopic observables of the population activity and also discuss the methods for rigorous mathematical proofs. Finally, I will discuss the most important features of relevant dynamical states observed in real data from neural networks in the visual cortex.

Lenka Zdeborová (Saclay):

Inference on networks via cavity method and message passing

In these lectures we will cover basics about connection between statistical mechanics and statistical inference. We will derive and show how to use methods developed in statistical physics – such as the cavity method and the associated message passing algorithms for models on statistical inference. We will present examples on sparse graphs but also on dense ones related to low rank matrix estimation. If time permits we will cover simple version of a proof of the replica formula for matrix estimation.

The participants' talks:

George Argyris

Suppose that we have a set that contains n dots on a paper. We choose two of them randomly and then perform the following coin-toss experiment: if "heads" we connect the two dots, and if "tails", we don't. Then, we repeat this experiment for each available pair of nodes on the paper. Afterwards, we perform a dice experiment for each available pair of dots, on a second piece of paper that is filled with dots, until we run out of pairs: if the result of a throw is "1", we connect the dots, otherwise we don't. What we will have on these 2 papers after the end of the experiments are random networks. Suppose, now, that we have an observed network. Is it possible to determine whether it came from a random experiment like the one we described? In this presentation, an index/scale that checks/quantifies if a network originates from such a random experiment will be mathematically introduced. That index is able to separate two highly probabilistic situations, as well as identify a kind of trend in stock market indices.

Suman Chakraborty

Weighted exponential random graph models: scope and large network limits

We study models of weighted exponential random graphs in the large network limit. These models have recently been proposed to model weighted network data arising from a host of applications including socio-econometric data such as migration flows and neuroscience. We present limiting results for the structure of these models as $n \to \infty$. We also present sufficient conditions for continuity of functionals in the specification of the model including conditions on node covariates.

Carolina Fransson

SIR epidemics on random graphs with clustering

One of the most important factors that determine the fate of an outbreak of an infectious disease is the contact patterns in the population. The frequency and duration of the contacts between individuals typically depend on the nature of their relationship. To incorporate social structure in an epidemic model the individuals of the population and the relationships between them may be represented by a graph; the nodes of the graph represent the individuals of the population and the edges represent social relationships. In this presentation, we consider SIR epidemics on random graphs generated by a version of the configuration model with clustering. SIR epidemics on graphs of this model have previously been investigated under the assumption of homogeneous infectivity within the population. We extend previous results by relaxing this assumption, allowing for variability in individual infectivity. By using a branching process approximation of the spread of the disease, expressions for the probability of a major outbreak and the expected final size is obtained.

Frank Marrs

Uncertainty quantification in network regression

Relational arrays represent interactions or associations between pairs of actors, often in varied contexts or over time. Such data appear as, for example, trade flows between countries, financial transactions between individuals, contact frequencies between school children in classrooms, and dynamic protein-protein interactions. We propose and evaluate a new class of parameter standard errors for models that represent elements of a relational array as a linear function of observable covariates. Uncertainty estimates for regression coefficients must account for both heterogeneity across actors and dependence arising from relations involving the same actor. Existing estimators of parameter standard errors that recognize such relational dependence rely on estimating extremely complex, heterogeneous structure across actors. Leveraging an exchangeability assumption, we derive parsimonious standard error estimators that pool information across actors and are substantially more accurate than existing estimators in a variety of settings. This exchangeability assumption is pervasive in network and array models in the statistics literature, but not previously considered when adjusting for dependence in a regression setting with relational data. We show that our estimator is consistent and demonstrate improvements in inference through simulation and a data set involving international trade.

Emily Muller

Determining the effects of social network evolution

The Stochastic Actor-Oriented Model is introduced as a means for estimating which network effects are significant determinants of social network evolution. Network effects include reciprocity, transitivity, homophily and more. Network evolution is assumed to be the result of individual friendship selection and selection is based on maximising utility relative to the effects and is dependent only on the current network (Markovian). The model is applied to a real-world social network and it's applicability to online social networks is explored.

Ido Nachum

Compression and learning

The notion of information compression scheme will be defined (an analog to sample compression scheme) and connections to PAC learning will be presented.

Alice Nanyanzi

Laplacian matrix of a network and applications

Many systems in real world are composed of components which are linked to each other to form complex systems. These systems include technological systems, ecological systems, social systems, biological systems among others. It is therefore of utmost importance to study the behaviour of these systems. This is achieved by representing the skeletons of complex systems by complex networks/large graphs whose properties are then studied. The Laplacian matrix is one of the matrices used to represent networks and it provides useful information on the properties of complex networks.This matrix has various applications in the study of networks. We consider two of the applications that is application to centrality measure and diffusion on networks.

Angelica Pachon

Scale-free behavior of networks with the copresence of preferential and uniform attachment rules

We introduce and study a random graph model which takes into account two different attachment rules: a preferential attachment mechanism that stresses the rich get richer system, and a uniform choice for the most recent nodes. This latter highlights a trend to select one of the last added nodes when no information is available. We obtain results on the degree distribution. Specifically we prove that this model exhibits an asymptotically power-law degree distribution when the recent nodes are taken as a fixed number. We also obtain numerical results for the case in which the recent nodes are given by a proportion of the total number of existing nodes. Our motivation is originated by real applications.

Snehal Shekatkar

Community structure in networks has become an intense topic of research since last decade. Most work in this area is however concentrated on quantifying the existence of communities in a principled way and finding out efficient algorithms to detect them. Consequently, a little attention has been given to find out an effect of the existence of communities on the function of the system represented by the network. In this work, we study an effect of the existence of communities on the network polarization: division of the nodes of the network in response to a yes-no type question. We find that the same network structure can give rise to very different polarized states depending upon the initial conditions. We thus argue that the information obtained by the methods of statistical inference may not be able to answer questions like why networks like Zachry karate club become polarized and eventually break.

WEEK ONE

	21 Aug	**22 Aug**	**23 Aug**	**24 Aug**	**25 Aug**
08:30 – 09:15	*Registration*				
09:15 – 09:30	*Welcome*				
9:30 – 11:00	Bhamidi	Polzehl	Bhamidi	Jahnel	Klimm
11:00 – 11:30	*coffee break*	*coffee break*	*coffee break*	*coffee break*	*coffee break*
11:30 – 13:00	Polzehl	Bhamidi	Jahnel	Klimm	Jahnel
13:00 – 14:20	*lunch break*	*lunch break*	*lunch*	*lunch break*	*lunch break*
14:20 – 16:00	Kyprianou, Mailler	Ex. Bhamidi		Ex. Jahnel	Ex. Klimm
16:00 – 16:30	*coffee break*	*coffee break*		*coffee break*	*coffee break*
16:30 – 18:00	Chak., Marrs	Arg., Muller		Shek., Nany.	
18:00	*Pizza welcome*				

WEEK TWO

	28 Aug	**29 Aug**	**30 Aug**	**31 Aug**	**1 Sep**
09:30 – 11:00	Stannat	Peixoto	Zdeborová	Zdeborová	Ex. Mörters
11:00 – 11:30	*coffee break*	*coffee break*	*coffee break*	*coffee break*	*coffee break*
11:30 – 13:00	Peixoto	Stannat	Zdeborová	Mörters	Mörters
13:00 – 14:20	*lunch break*	*lunch break*	*lunch*	*lunch break*	*lunch*
14:20 – 16:00	Ortgiese, Rogers	König, ???		Mörters	*departure*
16:00 – 16:30	*coffee break*	*coffee break*		*coffee break*	
16:30 – 18:00	Fransson, Nachum			Pachon	