

# BOOK OF ABSTRACTS

## Scientific committee:

- Prof. Ernst WIT (USI)
- Prof. Genise REINERT (Oxford)
- Prof. Maria Clelia DI SERIO (San Raffaele Milano)
- Prof. Steffen LAURITZEN (Copenhagen)
- Prof. Arnaldo FRIGESSI (Oslo)
- Prof. Goeran KAUEMANN (LMU)
- Dr. Claire GORMLEY (UDC)
- Prof. Vladimir Batagelj (Ljubljana)
- Dr. Veronica VINCOTTI (Brunel)
- Prof. Anuska Ferligoj (Ljubljana)

## Local organizing committee:

- Jesus M. Cortes Diaz (Local Chair). Ikerbasque, Biocruces Bizkaia & UPV/EHU
- Diego Rivera. Public University of Navarre
- Borja Camino Pontes. Biocruces Bizkaia & UPV/EHU
- Antonio Jimenez Marin. Biocruces Bizkaia & UPV/EHU
- Izaro Fernandez Iriondo. Biocruces Bizkaia & UPV/EHU
- Asier Erramuzpe. Edmond and Lily Safran & Biocruces Bizkaia
- Laiene Olabarrieta Landa. Public University of Navarre
- Begoña Landa Torre. Biocruces Bizkaia
- Libe Olabarrieta Landa. UPV/EHU
- Ernst Wit (Action Chair). Lugano University
- Gesine Reinert (Action Vice Chair). Oxford University

## Sponsors:

## Speakers:

1. Ernesto Estrada (University of Zaragoza, Spain)
2. Konstantin Klemm (IFISC (CSIC-UIB), Palma de Mallorca)
3. Alberto Roverato (University of Padova, Italy)
4. Lucia Paci (Università Cattolica del Sacro Cuore, Italy)
5. Paola Vicard (Università Roma Tre, Italy)
6. Monia Lupparelli (University of Florence, Italy)
7. Claudia Klüpperlberg (Technical University of Munich, Germany)
8. Gianpaolo Scalia Tomba (University of Rome Tor Vergata, Italy)
9. Cornelius Fritz (LMU, Germany)
10. Ernst Wit (Università della Svizzera italiana)
11. Rafiazka Hilman (Central European University, Austria)
12. Marjan Cugmas (University of Ljubljana, Slovenia)
13. Jesus Gomez-Gardenes (University of Zaragoza, Spain)
14. Miguel Angel Muñoz (University of Granada, Spain)
15. Jesus M. Cortes (Biocruces Bizkaia HRI, Spain)
16. Alessia Pini (Università Cattolica del Scaro Cuore, Italy)
17. Saverio Ranciati (University of Bologna, Italy)
18. Lasse Leskelä (Aalto University, Finland)
19. Rūta Užupytė (Vytautas Magnus University, Lithuania)
20. Fanny Villers (Sorbonne Université, France)
21. Silvia Fierascu (Central European University, Austria)
22. Sevag Kevork (LMU, Germany)
23. Antonietta Mira (Università dekkka Svizzera italiana, Italy)
24. Aryan Eftekhari (USI, Italy)
25. Gesine Reinert (University of Oxford, United Kingdom)
26. Diego Escribano Lopez (Universidad Carlos III, Spain)
27. Brieuc Lehmann (University of Oxford, United Kingdom)
28. Benjamin Sischka (Ludwig Maximilians Universität, Germany)
29. Vasiliki Koutra (King's College London, United Kingdom)
30. Luis Ospina-Forero (Alliance Manchester Business School, United Kingdom)
31. Costanza Tortù (IMT Lucca, Italy)
32. Pier Luigi Conti (Sapienza Università di Roma, Italy)

# Abstracts

# Communicability geometry of networks

Ernesto Estrada

(University of Zaragoza, Spain)

Starting from a motivation about diffusive processes on networks I will define the concept of network communicability. I will then prove how this matrix function induces a Euclidean embedding of a network. I will show how the communicability distance allows to understand some problems existing for the navigability of networks, such as brain networks, without global information. I will show several examples about urban traffic, community detection in gene-gene networks and brain damages.

Notes:

Notes:

# Structure of adaptive flow networks under load fluctuations

Konstantin Klemm

IFISC (CSIC-UIB)

Flow networks are crucial to the functioning of natural systems and technological infrastructures. In many scenarios, such as rivers or blood vessels, acyclic networks (i.e., trees) are the most efficient (optimal) solutions to a flow with given time-independent in- and out-flows at sources and sinks. Rather than global optimization, we here apply a local adaptation rule for the dynamics of the conductances: the change of a conductance is function only of the local pressure gradient. This reproduces the tree structures obtained under constant in- and out-flows. Imposing fluctuations of growing amplitude, however, renders the tree unstable and gives cyclic connections a non-zero conductivity in a transcritical bifurcation. The type of non-linearity in the adaptation rule controls the location of cyclic structures. For a super-linear dependence on pressure gradient, cycles first appear close to the tips ("capillaries"). We discuss these theoretical results in the light of real data of mammalian vasculature.

Notes:

Notes:

# Path weights in concentration graphs

Alberto Roverato

University of Padova, Italy

A graphical model provides a compact and efficient representation of the association structure of a multivariate distribution by means of a graph. Relevant features of the distribution are represented by vertices, edges and other higher-order graphical structures, such as cliques or paths. Typically, paths play a central role in these models because they determine the independence relationships among variables. However, while a theory of path coefficients is available in models for directed graphs, little has been investigated about the strength of the association represented by a path in an undirected graph. Essentially, it has been shown that the covariance between two variables can be decomposed into a sum of weights associated with each of the paths connecting the two variables in the corresponding concentration graph. In this context, we consider concentration graph models and provide an extensive analysis of the properties of path weights and their interpretation. More specifically, we give an interpretation of covariance weights through their factorisation into a partial covariance and an inflation factor. We then extend the covariance decomposition over the paths of an undirected graph to other measures of association, such as the marginal correlation coefficient and a quantity that we call the inflated correlation. We illustrate the usefulness of these findings by an application to a dietary intake network.

Notes:



Notes:

# Graphical model selection for air quality time series

Lucia Paci

Università Cattolica del Sacro Cuore, Italy

We propose an objective Bayes approach based on graphical models for learning dependencies among multiple air quality time series within the framework of Vector Autoregressive (VAR) models. Using a fractional Bayes factor approach, we obtain the marginal likelihood in closed form and construct an MCMC algorithm for Bayesian graphical model determination with limited computational burden. We apply our method to study the interactions between multiple air pollutants over the municipality of Milan (Italy).

Notes:

Notes:

# Statistical matching using Bayesian networks

Paola Vicard

Dipartimento di Economia, Università Roma Tre, Italy

Information for statistical analysis is frequently available in different micro data bases. Each data base contains some of the variables of interest. This is a serious drawback, when one is interested in the joint analysis of variables that are not jointly observed. Statistical matching aims at combining information obtained from different non-overlapping sample surveys, referred to the same target population. The main target is in constructing a complete synthetic data set where all the variables of interest are jointly observed. The lack of joint information on the variables of interest leads to uncertainty about the data generating model. In this paper we propose the use of Bayesian networks to deal with the statistical matching problem since they admit a recursive factorization of a joint distribution taking into account the conditional independence relations among all the variables of interest. Such a factorization is useful for the joint distribution estimation and for evaluating the statistical matching uncertainty in the multivariate context. In particular, the modularity of the graphical model allows to separately deal with: 1) subgraphs induced by nodes (variables) belonging to the same database; 2) subgraphs induced by variables observed on different samples. In other words, those factors generating uncertainty are separated from those directly estimable from the corresponding dataset. In this way also computation complexity is limited to subsets of variables. We also note that the conditional independence assumption (CIA), frequently used when dealing with statistical matching, here is not introduced. A simulation experiment has been performed in order both to evaluate the performance of the proposed methodology with and without auxiliary information, and to compare it, in terms of uncertainty reduction, with the saturated multinomial model. Results are all very promising showing a reduction of parameter uncertainty when Bayesian networks are used.

Notes:

Notes:

# Sliced-undirected and bi-directed graphical models

Monia Lupparelli

University of Florence, Italy

A novel class of sliced-graphical models is discussed for modelling the effect of an external risk factor on the dependence structure of a multivariate set of variables. The main aim is to provide a joint representation based on a single graph able to represent the sliced-distributions of a multivariate random vector given different levels of the risk factor. Conditional independence structures are explored by using sliced-undirected graphical models whereas marginal independence structures are given by sliced-bi-directed graph models. Markov properties specified for these two classes of graphical models show that they are compatible with chain graphs of different types known as LWF chain graph models and regression graph models, respectively.

Notes:

Notes:

# Conditional independence properties of recursive max-linear models

Claudia Klüppelberg

Technical University of Munich

Motivated by extreme value theory, max-linear graphical models have been recently introduced and studied as an alternative to the classical Gaussian or discrete distributions used in graphical modeling. We present max-linear models naturally in the framework of tropical geometry. This perspective allows us to shed light on some known results and to prove others with algebraic techniques. In particular, we give a complete description of conditional independence relations for max-linear recursive models.

Notes:



Notes:

# Analyzing a patient movement network in Norwegian hospitals

Gianpaolo Scalia Tomba

University of Rome Tor Vergata

One year of patient data for a large Norwegian region, containing the capital Oslo, has been analysed and various statistics related to patient career patterns and movements between wards and hospitals computed. A great heterogeneity of behaviours and preferential connections between wards and hospitals are found. Furthermore, the potential importance of patient movements for the spread of hospital infections will be discussed.

Notes:

Notes:

# A Separable Counting Processes-based Model for the Analysis of the International Arms Trade Network from 1950 to 2018

Cornelius Fritz

Department of Statistics LMU

We propose a novel tie-oriented model for the dynamic analysis of the international trade of major conventional weapons from the Stockholm International Peace Research Institute (SIPRI) between 1950 and 2018. The generating process consists of a multivariate Poisson counting process of event counts, that independently governs the formation and persistence of ties with two distinct intensity functions. The formation of a tie indicates an arms trade relation between two countries that was not present in the previous year, whereas a persisting tie occurs again. This characteristic is legitimized by the assumption that covariates have a different effect on repeated ties than on newly formed ones. Additionally, the complexity and extensive time frame of the observed arms trade network make time-varying effects necessary. Analogies between varying-coefficient models, time-to-event models, and counting processes facilitate the estimation of the parameters and account for more practical time-varying effects. The findings corroborate strongly differing and time-varying effects of endogenous and exogenous covariates on the processes of formation and persistence of arms trades. Especially during the end of the cold war period, the time-varying network effects display a fundamental structural change in the arms trade network.

Notes:

Notes:

# Estimating flow networks via systems of noisy differential equations

Ernst Wit  
Università della Svizzera italiana

In applied mathematics flow networks are typically thought of as non-linear systems of ordinary differential equations (ODE). These systems are a staple modelling tool in System Biology, Ecology or Biochemistry. The reasons of the generalized use of these models are their flexibility and ability to describe dynamical systems as flow networks. Despite the importance of ODE models, they have not been the focus of systematic statistical analysis until the last few years. In this work we propose a general approach to estimate the parameters of systems of differential equations measured with noise. Our methodology is based on the generalized Tychonov regularization where the differential system of equations is used as a penalty. Using this strategy, the solution of the ODE is not required which allows us to deal with large systems in which some components might be unobserved. We show the utility of the proposed method in a variety of scenarios ranging from the identification of gene regulatory networks to the estimation of the parameters of metabolic networks.

Notes:

Notes:

# A Robust Network Design for Sychromodal Transportation

Rafiazka Hilman

Central European University

Transportation network is highly connected in nature. It facilitates the mobility of passenger within public transportation network and the movement of goods within freight transportation network. Availability of transportation mode alternatives ranging from road network, waterway network, and railway network brings complexity yet flexibility in spatial and time dimension of such service. This research comes forward to tap this theoretical gap by offering an integrated network planning for transportation that involves synchronized multimodal transportation namely sychromodal transportation. This relates to the empirical goal on developing a contingency framework for mitigating and dealing with disruption at transfer points such as terminal and corridor where interactions between different transportation modes take place. One could imagine that a delay of container delivery caused by the technical failure of local barge system in Rotterdam may propagate throughout that particular transfer point or even affect the performance of wider network, to Duisburg for example. Number of questions stand on the center of this research: How susceptible is sychromodal transportation to disruption? How can we measure the robustness of such system? Is it possible to accommodate intermediate transfer at intermodal corridor if a disruption happens on a node (e.g. terminal, corridor) or at an edge (e.g. road/railways/waterways)? Does such design improve the effectiveness and efficiency of supply chains and passenger mobility? Three steps are taken into account. At first, sychromodal transportation is modeled as a multilayer network in which interdependency between layers exist. It is followed by searching and fitting robustness measures in sychromodal freight transportation network. Later, it examines to what extent sychromodal transportation network may deal with disruption such that cascading failure could be minimized by using network flow optimization model.

Notes:



Notes:

# The emergence of the global network structure of knowledge-flow networks

Marjan Cugmas

Faculty of Social Sciences, University of Ljubljana

Knowledge-flow networks describe how the knowledge is exchanged among the units. They are usually observed within different companies since understanding patterns and underlying mechanisms of exchanging knowledge among the employees is crucial to ensure the competitiveness of a company. The presentation addresses the evolution of a global network structure of the knowledge-flow networks by considering the selected local network mechanisms. The empirical data are analyzed by blockmodeling approach to identify the global network structure of the knowledge-flow networks. The data were collected among the employees in the international company. The results show that the global network structure is approaching to the hierarchical one in time. Once the global network structure is identified, the algorithm from the family of the network evolution models is used to test if the selected local network mechanisms can drive the global network structure towards the hierarchical one. The algorithm and the selected mechanisms are based on the theory proposed by Nebus (2006). By following his theory, the employees are considering the costs and the benefits of asking for advice to a given unit. It is confirmed, by the Monte Carlo simulations, that the hierarchical global network structure can emerge as a consequence of the mechanisms which are related to the hierarchical position of the units, tenure of the units, popularity level of the units, homophily of the shared partners and distance between the ego and the alter.

Notes:

Notes:

# Data-driven models to evaluate the impact of human activity in the spread of diseases

Jesus Gomez-Gardenes

University of Zaragoza

Agent-based simulations provide with a detailed description of disease spreading, as they constitute the ideal benchmark to incorporate detailed information about human activity and epidemiological data about the spread of pathogens. However, agent-based simulations fail in giving insights about the role played by each ingredient in the onset of epidemic states. Here we will show that it is possible to construct metapopulation models that, apart from being amenable for their mathematical analysis, can accommodate real data about human and contagion patterns. We will apply this formalism to the treatment of vector-borne disease to derive a risk metric for each subpopulation. We will show the agreement of this measure with the real incidence of Dengue in the city of Cali (Colombia) where this disease is endemic.

Notes:

Notes:

# Critical dynamics and Griffiths phases in the brain

Miguel Angel Muñoz

University of Granada

In this talk I will briefly review the idea or hypothesis that the human brain might operate in a dynamical regime close the edge of a phase transition. Such a phase transition separates “physical phases” with very different dynamical properties. Lying t the very edge of the phase transition could provide the system with exceptional functional capabilities.

Notes:

Notes:

# Multi-scale Brain Hierarchical Networks

Jesus M Cortes

Biocruces Bizkaia HRI

Despite growing theories and knowledge in the relation between structure and function in network modeling, most of the studied scenarios are not sufficient for explaining brain data, where the multi-scale nature of both structure and function makes their matching yet unknown. Here, we adopt a systems approach that uses modular hierarchical clustering to combine brain anatomical networks together with functional ones (Diez et al., *Sci. Rep.*, 2015), resulting into a common framework. I will discuss several examples of clinical research where the use of this strategy has succeed. First, I will show how brain networks reorganize and re-connect to new regions after traumatic brain injury in a pediatric population (Diez et al, *Netw. Neurosci*, 2017). Second, I will show new network biomarkers in ageing (Bonifazi et al, *Hum. Brain Map.* 2018). Third, I will show the presence of progressive alterations in brain networks across severity stages in Alzheimer's disease (Rasero et al , *Front. Aging Neurosci.*, 2017). Finally, how redundancy and synergy reveal new informational role in brain networks (Camino-Pontes et al., *Entropy*, 2018).

Notes:



Notes:

# Statistical comparison of two groups of networks: an application to brain connectivity of patients with autism

Alessia Pini

Università Cattolica del Sacro Cuore

Networks have been extensively used for representing the human brain and for studying its structure and functions. The human brain is represented as a network whose knots are different location in the cortex and whose edges are the connections between such locations. In this framework, the investigation of neurological disorders can be performed by comparing samples of brain networks between healthy and non-healthy subjects, or between subjects affected by different types of disorders. This problem falls under the object-oriented data analysis (OODA) paradigm: for each subject, the object of the statistical analysis is not a number nor a vector, but a complex object, which is in this case a whole network. The aim of this work is to present a method able to test for differences between two populations of networks. While it is interesting to determine whether the two samples of networks have been generated by two different distributions, it is key, for understanding the biological mechanisms involved in the pathology, to localize the differences on the brain network. In particular, it is of great importance to localize the knots and edges that present statistically significant differences between the two groups. Motivated by this question, we propose a non-parametric finite-sample exact statistical test that allows to test for differences in connectivity within and between pre-specified regions of interest inside the brain network. The test is applied to brain structural connectivity networks obtained from electroencephalography, to differentiate children with non-syndromic autism from children with both autism and tuberous sclerosis complex.

Notes:

Notes:

# Fused graphical lasso with edge symmetries: an application to resting state fMRI

Saverio Ranciati

Department of Statistical Sciences - University of Bologna

Neuroscientists nowadays are interested in discovering patterns of activations of specific areas of the human brain, known to be responsible for many behaviours and biological activities. From a statistical perspective, we could phrase this problem as finding a network of dependencies in the data coming from neuroimaging experiments. The aim of this work is to present a graphical lasso procedure, with a novel focus on highlighting symmetric activations of the left and right hemisphere. This is achieved by using a penalization term that encourages symmetric values inside two blocks of the same concentration matrix.

Notes:

Notes:

# Statistical inference of dynamic stochastic block models

Lasse Leskelä

Aalto University

Dynamic stochastic block models with static node attributes are used to model time-varying networks where pairwise interactions between nodes are conditionally independent given node attributes. The node attributes induce a latent community structure. I will discuss the problem of inferring the community structure and the dynamic interaction parameters based on observing a correlated graph time series, and present our latest findings on how much the observing multiple correlated graph samples helps in improving the accuracy of community detection compared to standard stochastic block models based on a single graph sample. The talk is based on joint work with Konstantin Avrachenkov and Maximilien Drevetton (Inria Sophia Antipolis, France).

Notes:

Notes:

# A framework for event-based network modeling and analysis

Rūta Užupytė

Vytautas Magnus University

Temporal networks are increasingly being used to model the interactions of complex systems. However, most of the studies require the temporal aggregation of events into discrete time steps to perform analysis. Nowadays, the availability of time-stamped event data is increasing due to the digitalized archives and electronic devices recording the exact time of online and offline communications. Hence, new tools are required in order to describe and analyze the behavior of networks presenting event-based data. The aim of this research is to analyze possible approaches for the representation of temporal event graph, taking into account the underlying network structure, parameters responsible for the network dynamics, characteristics and patterns describing communicative behaviour, etc.

Notes:



Notes:

# Graph inference with clustering and false discovery rate control

Fanny Villers

Sorbonne Université

Networks analysis is widely used to model and describe interactions between entities. In many situations, the practitioner does not directly observe the interaction graph, but only a noisy version of it. This is for example the case of similarity graphs or when edges represent correlations between the nodes. To recover the underlying graph, ad hoc methods such as thresholding rules are often used, which can cause a degradation of the topology of the network and may lead to wrong interpretations, especially when combined with cluster analysis. In this work, we introduce a noisy version of the stochastic block model and propose to recover the underlying graph with a false discovery rate (FDR) guarantee, that is, by controlling the average proportion of errors among the edges declared as significant. A novelty of our method is that the graph and the clustering are inferred simultaneously, so that the two types of inference can help each other. Our methodology relies upon Bayesian multiple testing techniques with parameters estimated by a suitable variational EM algorithm. The performances of our method are supported both by theoretical findings and by numerical experiments.

Notes:

Notes:

# **Pathways to Capturing the State: Comparative Analysis of Corruption Networks in Public Procurement across European Countries**

Silvia Fierascu

Central European University

The study of high-level corruption is not new. But the study of corruption by pursuing the distribution of public money as quantifiable networks based on large and objective data is a new approach. This project aims to analyze big data on public procurement networks of public institutions and private organizations across 31 European countries, between 2009 and 2014. The goal is to reveal the relational building blocks of corruption networks in public procurement, and to compare levels of state capture, varieties of state capture (business versus political capture) and geographical patterns of state capture across these countries. By mapping and analyzing more than 1 million public procurement contracts linking an estimated of more than 300 billion euros into corruption networks, I will be able to make explicit and test the mechanisms at work behind the formation, evolution and dynamics of corruption networks across procurement markets, countries and over time.

Notes:

Notes:

# Iterative Estimation for ERGMs with Nodal Random Effects

Sevag Kevork

COSTNET

The presence of unobserved node specific heterogeneity in Exponential Random Graph Models (ERGM) is a general concern which causes instability in estimation. We extend the well-known ERGM by including random node effects to account for unobserved heterogeneity in the network. This leads to an ERGM with random structure on the coefficients. Estimation is carried out by combining approximate penalized pseudo-likelihood estimation for the random effects with maximum likelihood estimation for the remaining parameters in the model. This allows to fit nodal heterogeneity effects even for large scale networks.

Notes:

Notes:

# Parallelized High-Dimensional Sparse Inverse Covariance Matrix Estimation

Aryan Eftekhari

USI

The estimation of large sparse inverse covariance matrices is an ubiquitous statistical problem in many application areas such as mathematical finance or geology or many others. Numerical approaches typically rely on the maximum likelihood estimation or its negative log-likelihood function. When the Gaussian mean random field is expected to be sparse, regularization techniques which add a sparsity prior have become popular to address this issue. This problem is particularly challenging as the required computational resources increase superlinearly with the number of random variables. We introduce a performant and scalable algorithm [1,2] which builds on the current advancements of second-order methods. The routine leverages the intrinsic parallelism in the linear algebra operations and exploits the underlying sparsity of the problem. The developed SQUIC framework is used to approximate the sparse inverse covariance matrix for datasets with up to 10 million random variables.

Notes:



Notes:

# Approximating exponential random graphs

Gesine Reinert

University of Oxford

We provide a general bound on the Wasserstein distance between two arbitrary distributions of sequences of Bernoulli random variables.

The result is applied to estimate, with explicit error, expectations of functions of random vectors for exponential random graphs in "high temperature" regimes.

This is joint work with Nathan Ross, University of Melbourne.

Notes:

Notes:

# Social networks structure using exponential random graphs

Diego Escribano Gómez

Universidad Carlos III Madrid

Humans organize their social relationships in networks with a characteristic size and layered structure. An individual is able to manage about 150 relations, which are classified into a set of inclusive layers. The size of each layer is inversely proportional to the emotional intensity of their associated links and exhibits a constant scaling ratio. No method has been previously proposed to explain social networks structure including these evidences. Here we show that the framework of human relations can be explained just by taking into account three factors: the limited cognitive capacity, the personal layered structure and the reciprocity of the links. For that purpose, we use exponential random graphs to develop a pair interaction model capable to fit experimental data and characterize the global system in equilibrium. Therefore, our model allows to structure a network from personal features and provides a vision of social communities different to the classical one.

Notes:

Notes:

# Inferring differences between groups of networks using Bayesian exponential random graph models

Brieuc Lehmann

University of Oxford

The goal of many neuroimaging studies is to better understand how the functional connectivity structure of the brain changes with a given phenotype such as age. Functional connectivity can be characterised as a network, with nodes corresponding to brain regions and edges corresponding to statistical dependencies between the respective regional time series of activity. A typical neuroimaging dataset will thus consist of one or more networks for each individual in the study. Most statistical network models, however, were originally proposed to describe a single underlying relational structure such as friendships between individuals or hyperlinks between web pages. As a result, the development of these models has largely been restricted to the single network case. While one could in principle fit a single network model to each individual separately, it is not always straightforward to combine these individual results into a single group result.

We propose a multilevel framework for populations of networks based on exponential random graph models. By pooling information across the individual networks, this framework provides a principled approach to characterise the relational structure for an entire population. We use the framework to assess group-level variations in functional connectivity, providing a method for the inference of differences in the topological structure between groups of networks.

Notes:

Notes:

# Estimating Stochastic Block Models by using their Graphon Representations

Benjamin Sischka

Ludwig-Maximilians-Universität München

We propose a new Stochastic Block Model estimation by making use of the representation as a (discrete) Graphon Model. Therefore, our previously developed EM type graphon estimation routine can be adapted [1]. Applying the Gibbs sampler yields information about the repositioning of the nodes. In case of a community structure the posterior mean values will be arranged cluster-wise. Based on the adjacency matrix and the node positioning, we are using linear B-splines to achieve a (continuous) graphon estimate. This iterative estimation routine is able to capture the underlying community structure, except for the extent of the transition regions between the communities. Quantile adjustment referring to the predicted positioning of the nodes is used to compress the disproportionated transition regions to their appropriate size. This procedure yields a continuous approximation of the discrete structure. However, if required, selective cut points can easily be derived and communities can be categorized.

Notes:



Notes:

# Design of experiments with crossed blocking structures and network effects for use in agricultural trials

Vasiliki Koutra

King's College London

In this talk we propose a new method of constructing efficient designs with complex blocking structures and network effects and we illustrate it in an agricultural field experiment. The idea is to view the potential interference among plots that can result from the spatial structure or farmer operations as a network structure. We consider a field trial running at Rothamsted Experimental Station and we obtain different row-column designs accounting for the neighbour effects, implementing an interchange algorithm. We provide a comparison of various optimal designs under different models, including the commonly used designs in such situations. We show that when ignoring the neighbour structure the designs are poor and can lead to inaccurate estimates of the treatment effects' differences and false conclusions.

Notes:

Notes:

# Challenges on the estimation of Sustainable Goal Networks

Luis Ospina-Forero

Alliance Manchester Business School - University of Manchester

As the United Nations 2030 Agenda has progressed, it has been pointed out that, in order to truly achieve sustainable development, it is necessary to understand how the multiple dimensions of sustainable development interact with one another. In other words, a multidimensional view of development requires well-defined procedures to quantify and operationalize networks of interdependencies between different goals (or their indicators).

This work considers the use different causal network-estimation methods with the aim of identifying the most suitable procedures to build networks of sustainable development indicators. For this reason, we explore their underlying assumptions, strengths and limitations. Broadly speaking, we classify current available methods into five families: correlation thresholding, Granger causality, chordal information filtering, statistical structure learning and physics-inspired methods. Interestingly, only a handful of methods can be employed with empirical development-indicator data. In this study we firstly perform a simple comparative study using those methods, thus illustrating their potential for estimating sustainable development goal networks. And secondly, we provide a list of challenges of sustainable development goal networks that portrays the reasons why most structural causal network inference methods fail or struggle at providing causal estimations of these sustainable development goals, thus augmenting the need for more suitable causal network estimation methodologies.

Notes:

Notes:

# The Causal Effect of Immigration Policies: Modelling Network Interference with a Multivalued Treatment Exposure

Costanza Tortù

IMT Lucca

The aim of this work is to analyze the impact of immigration policies on crime rates and inequality. Starting from the IMPIC dataset, reporting all immigration policies conducted by the OECD countries between 1980 and 2010, we use IRT models to identify latent factors measuring the attitude of a given country-year toward immigration with respect to the major political issues. This defines immigration policies as a multivariate treatment and causal effects can be expressed as comparisons of crime rates and inequality under different levels of this treatment. Estimation of these causal effects is further complicated by the plausible presence of interference between countries: immigration policies of one country may influence the outcomes of other countries. The extent to which each country can be influenced by another country depends on an Interference Compound Index (ICI), which depends on geographical proximity, cultural proximity and trading volume. Each country is then subject to national immigration policies and exposed to a the international policies of other countries, weighted by the ICI. The statistical framework builds on the growing literature on causal inference in the presence of network interference and the proposed estimator capitalizes on the use of a joint propensity score to adjust for covariate imbalance across national and international treatment levels.

Notes:

Notes:

# Sampling design issues in observational studies, with applications to the estimation of the distribution function of potential outcomes.

Pier Luigi Conti

Sapienza Università di Roma

The evaluation of the possible effects of a treatment on an outcome plays a preeminent role in bio-statistics, as well as in micro-econometrics, mainly in policy evaluation.

To be concrete, let  $Y$  be an outcome of interest, observed on  $N$  subjects. Some of them are treated by an appropriate treatment, whilst other subjects may be considered as untreated, and can be considered as a control group. Let  $T$  denote the “treatment variable”, taking (conventional) values  $0, 1, \dots, K$ . In our convention,  $T = k$  means “level  $k$  of treatment is applied” ( $k = 0, 1, \dots, K$ ), and in particular  $T = 0$  means “absence of treatment”. Let further  $Y(k)$  be the (potential) outcome due to receiving treatment at level  $k$ . The observed outcome  $Y$  is then equal to  $Y(k)$  whenever  $T = k$ .

The main source of trouble is that receiving a treatment level is not a “purely random” event, as in experimental framework. On the contrary, there could be relevant differences between treated and untreated subjects, due to the presence of confounding covariates. In the sequel, we will denote by  $X$  the (random) vector of relevant covariates, that is assumed to be observed. In the sampling terminology, the main problem is that the sampling design assigning subjects to treatment levels is essentially unknown. In the present talk, a modelization of the sampling design as a Poisson design with unknown first order inclusion probabilities is considered, taking into account, in particular, the possible dependence among covariates. Estimation of such probabilities (that, in the case  $K = 1$ , are called propensity scores) is discussed. Then, a Hájek estimator of the distribution function of  $Y(k)$ s is proposed, and its asymptotic distribution is studied. Applications to various non-parametric tests are finally considered.

Notes:



Notes: