Action Conference and Working Group Meetings

Statistical Network Science

COSTNET17

IFISC, University of the Balearic Islands
Palma de Mallorca, 25-27 October, 2017

Book of Abstracts

last update: October 18, 2017

Perceptions in Social Network Analysis:
Statistical Modelling of Three-way Data

1

M. A. J. van Duijn

Department of Sociology, University of Groningen
The Netherlands

Social network data are usually collected as self-report dyadic data: actors assess the presence, absence or strength of one or more relationships with other actors in a well-defined group. The only information available about agreement on the ties comes from reciprocity in directed networks. In some applications, however, dyads are reported by others than the actors involved in the data. Well-known examples are cognitive social structures where actors in the network not only report about their own ties, but also assess the ties between all other actors. The analysis of the resulting triadic data is not straightforward, as there are few models for triadic data, and many ways for reducing the data to dyadic data. The case of ego-centric data where ego reports on the ties between the alters in his/her social network may be one of the simpler cases to deal with. In other applications, self-report data can be combined with information from external sources on the social network, for instance when parents and teachers assess the positive or negative relationships between children, or with information on other ties between the actors in the network, leading to a multiplex social network analysis. In the talk I will present models and applications on how to deal with perceptions in social network analysis and would like to discuss their problems and possible solutions.

Regression graph models
for binary non-independent outcomes

Monia Lupparelli

Department of Statistical Sciences, University of Bologna, IT

Regression models represent a common framework for modelling the effect of risk factors on a suitable set of outcomes. When the endpoint of interest in statistical analysis involves non-independent outcomes, a multivariate regression framework is required in order to account for the outcome dependence. A graphical approach for modelling the joint distribution of outcomes and explanatory variables is illustrated based on a sequence of non-independent regressions; see [4]. In particular, the focus is on the special case of binary outcomes which rises serious challenges because the outcome dependence needs to be modelled including a sequence of marginal regressions. In this context it is useful to recall the notion of product outcomes representing the endpoint of interest in marginal regressions and the decomposition of the risk factor effect on product outcomes into the intrinsic and the extrinsic component providing useful insight on the model interpretation; see [2]. In the last decade several parameterizations for discrete regression graphs have been developed, however the log-mean linear parameterization of [3] is shown to be appealing in case of binary outcomes because risk factor effects of interest can be derived by model parameters. An algorithm for maximum likelihood estimates inspired on [1] is illustrated. The algorithm is general and not model dependent, then it can be employed for fitting discrete regression graph models whichever parameterization is used.

1 Lang, J. B. (1996). Maximum likelihood methods for a generalized class of log-linear models. *The Annals of Statistics*, **24**, 726-752.

2 Lupparelli, M. and Mattei, A. (2017). Causal inference for binary non-independent outcomes (*submitted*).

3 Lupparelli, M. and Roverato, A. (2017). Log-mean linear regression models for binary responses with an application to multimorbidity. *Journal of the Royal Statistical Society, Series C*, **66**(2), 227-252.

4 Wermuth, N. and Sadeghi, K.(2012). Sequences of regressions and their independences. *Test*, **21**, 215-252.

Nonparametric weighted stochastic block models

TIAGO P. PEIXOTO

Department of Mathematical Sciences, University of Bath, UK

We present a Bayesian formulation of weighted stochastic block models that can be used to infer the large-scale modular structure of weighted networks, including their hierarchical organization. Our method is nonparametric, and thus does not require the prior knowledge of the number of groups or other dimensions of the model, which are instead inferred from data. We give a comprehensive treatment of different kinds of edge weights (i.e. continuous or discrete, signed or unsigned, bounded or unbounded), as well as arbitrary weight transformations, and describe an unsupervised model selection approach to choose the best network description. We illustrate the application of our method to a variety of empirical weighted networks, such as global migrations, voting patterns in congress, and neural connections in the human brain.

Gene Trees, Species Trees, Reconciliation Maps, and Phylogenenies.

4

Peter F. Stadler[1,2,3,4,5,6]

[1] Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center of Bioinformatics, Leipzig University, D-04107 Leipzig, Germany
[2] Max Planck Institute for Mathematics in the Sciences, D-04103 Leipzig, Germany
[3] Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany
[4] Institute for Theoretical Chemistry, University of Vienna, A-1090 Vienna, Austria
[5] Center for Non-Coding RNA in Technology and Health, University of Copenhagen, 1870 Frediksberg C, Denmark
[6] Santa Fe Institute, Santa Fe, NM 87501

Sequence-based phylogenetic approaches heavily rely on initial data sets to be composed of orthologous sequences only. Paralogs are usually treated as a dangerous nuisance that has to be detected and removed. Recent advances in mathematical phylogenetics, however, have indicated that gene duplications can also convey meaningful phylogenetic information provided orthologs and paralogs can be distinguished with a degree of certainty.

A key observation is that orthology and paralogy relations have co-graph structure. This makes it possible in practice to correct initial estimates. The co-trees equivalent to co-graphs now are incompletely resolved versions of the gene trees with events (duplications or speciations) explicitly annotated. Since these event-labeled gene trees are embedded in the species tree, they encode constraints. In a phylogenomic setting, these constraints are sufficiently informative to completely specify the species tree.

In the presence of horizontal gene transfer (HGT) the situation becomes more complicated since HGT is inherent non-symmetric: one copy of the gene is transferred to a distinct branch of the tree, while the other one remains in its ancestral context. This situation is captured by non-symmetric relations and trees that are either though of as oriented or edge-labeled. We recently obtained complete characterization of these classes of xenology relations and devised algorithms to reconstruct the tree information contained in them, providing direct access to many important aspects of the history of large gene families.

Unveiling untapped potential in cities:
From taxi sharing to mobility space inequality

Michael Szell[1,2,3,4]

[1] Center for Network Science, Central European University, Hungary
[2] Centre for Social Science, Hungarian Academy of Sciences, Hungary
[3] Center for Complex Network Research, Northeastern University, USA
[4] Complexity Science Hub Vienna, Austria

We introduce the notion of shareability network, allowing us to model the collective benefits of sharing rides as a function of passenger inconvenience, and to efficiently compute optimal sharing strategies on massive datasets. We apply this framework to millions of taxi trips in New York City, showing that with low passenger discomfort, cumulative trip length can be cut by 40%. This benefit comes with reductions in service cost, emissions, and split fares. Shareability as a function of trip density saturates fast, suggesting effectiveness of the sharing system also in cities with much sparser taxi fleets or when willingness to share is low [1]. Applying the framework to diverse cities, we find a collapse of shareability curves onto a single, universal curve. We explain this scaling law with a simple model that predicts the potential for ride sharing in any city, using a few basic urban quantities [2]. Finally, we present "What the Street!?", an online platform for the interactive exploration of city-wide mobility spaces. The aim of What the Street!? is to facilitate the intuitive exploration of (wasted) mobility space in cities, exploring to which extent space is distributed unevenly between different modes of transportation. We demonstrate how this data visualization of re-ordered city spaces can inform stakeholders and the public about large-scale reductions of parking spaces in scenarios of wide-spread car-sharing [3].

1. P. Santi, G. Resta, M. Szell, S. Sobolevsky, S. Strogatz, C. Ratti. *Quantifying the benefits of vehicle pooling with shareability networks*, PNAS 111(37), 13290-13294 (2014).

2. R. Tachet, O. Sagarra, P. Santi, G. Resta, M. Szell, S. Strogatz, C. Ratti. *Scaling law of urban ride sharing*, Scientific Reports 7, 42868 (2017).

3. M. Szell. *Crowdsourced quantification and visualization of urban mobility space inequality*, Working paper (2017).

Improving coextinction models to simulate consequences of pollinator loss on plant communities

Anna Traveset[1], Cristina Tur[1], Víctor M. Eguíluz[2]

[1] Institut Mediterrani d'Estudis Avanats IMEDEA (CSIC-UIB), Global Change Research Group, C/Miquel Marqus 21, E07190-Esporles, Mallorca, Spain.
[2] Instituto de Fsica Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), E07122, Palma de Mallorca, Spain.

Coextinction models are useful to understand community robustness to species loss and resilience to disturbances. We simulated pollinator extinctions in pollination networks by using a hybrid model that combined a recently developed stochastic coextinction model (SCM) for plant extinctions and a topological model (TCM) for animal extinctions. Our model accounted for variation in interaction strengths and included empirical estimates of plant dependence on pollinators to set seeds. The stochastic nature of such model allowed us determining plant survival to single (and multiple) extinction events, and identifying which pollinators (keystone species) were more likely to trigger secondary extinctions. Consistently across three different pollinator removal sequences, plant robustness was lower than in a pure TCM, and plant survival was more determined by dependence on the mutualism than by interaction strength. As expected, highly connected and dependent plants were the most sensitive to pollinator loss and collapsed faster in extinction cascades. We predict that the relationship between dependence and plant connectivity is crucial to determine network robustness to interaction loss. Finally, we showed that honeybees and several beetles were keystone species in our communities. This information is of great value to foresee consequences of pollinator losses facing current global change and to identify target species for effective conservation.

ℓ_1-penalized Gaussian graphical model for left-censored data

ANTONINO ABBRUZZO[1], LUIGI AUGUGLIARO[1], VERONICA VINCIOTTI[2]

[1] Department SEAS, University of Palermo, Italy
[2] Department of Mathematics, Brunel University, London.

Penalized Gaussian graphical models (pGGM) have been extensively used to estimate conditional independence structures among a large set of multivariate random variable in several fields of science, e.g. biology. However, in the biomarker detection study, due to specific limits of quantification of the measurement technologies, some biomarkers are difficult to measure and/or are measured with less than optimally refined laboratory methods. This data generating mechanism produces continuous data with left-censure which makes the pGGM theoretically unfounded.

We propose an extension of the ℓ_1 penalized Gaussian graphical model, which allows making parameter estimation and model selection, simultaneously. This extension is motivated by an EM-like algorithm which provides an accurate estimate of the conditional independence graph at a reasonable computational cost.

Ageing and multiplexity in the voter model

O. ARTIME[1], J. FERNÁNDEZ-GRACIA[1], J. RAMASCO[1], M. SAN MIGUEL[1]

[1] IFISC (CSIC-UIB), Campus UIB, 07122 Palma de Mallorca, Spain.

The voter model rules are simple, but they lead to non-trivial dynamics. Inspired by the temporal inhomogeneities found in human interactions, one can introduce ageing in the agents: the probability to update their state decreases with the time elapsed since the last change. This modified dynamics induces an approach to consensus via coarsening in complex networks. In this work, we investigate how a multilayer structure affects the dynamics of the ageing voter model. The system is studied as a function of the fraction of nodes q sharing states across layers. We find that the dynamics of the system is rich, suffering a notable change at an intermediate value q^*. Above it, the voter model always orders via coarsening to an absorbing configuration, no matter the dimensionality of the network. Below q^*, the system has a finite probability to end up into dynamical traps associated to a spontaneous symmetry breaking. Further details can be found in reference [1].

1 O. Artime *et al.*, "Joint effect of ageing and multilayer structure prevents ordering in the voter model" , Scientific reports 7 (2017).

Dynamics on Networks: Competition between Temporal and Topological Correlations

Oriol Artime[1], Jose J. Ramasco[1], Maxi San Miguel[1]

[1] IFISC (CSIC-UIB), Campus UIB, 07122 Palma de Mallorca, Spain.

In real-world communication data one finds burstiness, as well as several correlations in the process of link activation: memory effects imply temporal correlations, but also the existence of communities in the network may mediate the activation patterns of internal an external links. In this poster we show the effect of the competition of topological and temporal correlations in link activation and how they affect the dynamics of systems running on the network. Interestingly, both types of correlations by separate have opposite effects: one (topological) delays the dynamics of processes on the network, while the other (temporal) accelerates it. When they occur together, our results show that the direction and intensity of the final outcome depends on the competition in a non trivial way.

1 O. Artime, J. Ramasco, & M. San Miguel, "Dynamics on networks: competition of temporal and topological correlations" , Scientific Reports 7, 41627 (2017).

Detecting random walks on graphs with heterogeneous sensors

Dragana Bajović[1], José M. F. Moura[2], Dejan Vukobratović[1]

[1] Faculty of Technical Sciences, University of Novi Sad, Serbia
[2] Carnegie Mellon University, Pittsburgh, PA, USA

We consider the problem of computing error exponent for detection of a random walk on a graph. When visited by the walk, graph nodes observe signals of elevated mean, possibly different at different nodes; outside of the path of the walk nodes measure only noise. Assuming the Neyman-Pearson setting, our goal is to characterize detection performance by computing the error exponent for the probability of miss, under a constraint on the probability of false alarm, when the observation interval grows large. Since exact computation of the error exponent is known to be a difficult problem (equivalent to computation of the Lyapunov exponent) we approximate its value by finding a tractable lower bound. The bound reveals an interesting condition: the walk is detectable whenever the entropy of the walk is smaller than one half of the expected signal to noise ratio. Finally, we show by a sequence of judicious problem reformulations that computing the bound reduces to solving a convex optimization problem, which is a result in its own right.

Scaling in the recovery of cities during special events

11

A.Bassolas[1], R. Gallotti[1], F. Lamanna, M. Lenormand[2] and JJ. Ramasco[1]

[1] IFISC, Campus Universitat de les Illes Balears, Palma de Mallorca E-07122, Spain
[2] Irstea, UMR TETIS, 500 rue Franois Breton, Montpellier, France

Public transportation system must cope with increased demand in exceptional crowd gathering events such as concerts, football matches or other localized events. Here we study the emergence of delays and the collapse of a public transport network under situation of stress. In this work we use a simplified model to simulate the use of the public transport system in a city and test the performance of the network under different routing protocols. We then study the recovery of cities from huge events, introducing perturbations of different sizes in localized places. Depending on the routing protocols used we observe a different recovery of the city. We focus on understanding the recovery times with the amount of individuals and the place of the perturbation. We find that while the local dimension of the network dominates the scaling with the number of individuals, the recovery time of a zone depends on the time that an individual needs to walk to find a non congested stop.

Describing network evolution using probabilistic inductive classes

12

Vladimir Batagelj

IMFM, Ljubljana, Slovenia
vladimir.batagelj@fmf.uni-lj.si

An approach to describe the evolution of networks is to use probabilistic inductive classes [1]: starting with a selected basic network we apply a random sequence of rules that transform the network.

We will present basic ideas of the approach and illustrate it with some examples.

1 Nataa Kejar, Zoran Nikoloski, Vladimir Batagelj: Probabilistic Inductive Classes of Graphs. Journal of Mathematical Sociology, 32:85109, 2008.

The CostNet website

Vladimir Batagelj

IMFM, Ljubljana, Slovenia
vladimir.batagelj@fmf.uni-lj.si

The CostNet website http://costnet.webhosting.rug.nl/dokuwiki is based on a wiki. This allows project members not only to read its content but also to contribute to it. We will present the current structure of the website (About, Contacts, Members, Working groups, STSM, News, Events, Publications, Galleries, Data, Internal) and explain how a project member can contribute - changing text in a page, basic elements of the text formating (headings, lists, tables, figures, etc.), creating a new page, uploading files, etc.

Dynamic Chain Graph Models for Time Series Mixed Data

Pariya Behrouzi[1,2], Ernst C. Wit[1]

[1] Faculty of Mathematics and Natural Sciences, University of Groningen, NL
[2] Mathematical and Statistical Method, Wageningen University, NL

This paper introduces a sparse dynamic chain graph models for network inference in high dimensional non-Gaussian time series data. The proposed method parametrized by a precision matrix that encodes the intra time-slice conditional independences among variables at a fixed time point, and an autoregressive coefficient that contains dynamic conditional independences interactions among time series components across consecutive time steps. The estimation of the parameters in the proposed method relies on Gaussian copula graphical models and Bayesian networks under penalized expectation-maximization (EM) algorithm framework. In this paper, we use an efficient coordinate descent algorithm to optimize the penalized log-likelihood with the smoothly clipped absolute deviation penalty. We demonstrate our approach on simulated and genomic datasets. We have implemented the method as a general purpose package in R.

Thresholding and random network generation based on uncertain networks

15

<u>L V Bozhilova</u>[1], A V Whitmore[2], J Wray[2], C M Deane[1], G Reinert[1]

[1] Department of Statistics, University of Oxford, Oxford, UK
[2] e-Therapeutics, Oxford, UK

In order to construct protein interaction networks (PINs), inhomogeneous interaction data can be combined and scored based on its reliability. This results in uncertain networks [1]. Given an uncertain network, it is of interest to estimate structural properties of the unobserved underlying true network, i.e. of the biologically relevant cellular state. In this work we compare an uncertain network representing the yeast (*S. cerevisiae*) interactome to a synthetic uncertain PIN via thresholding and random network generation. We further illustrate some of the difficulties associated with estimating true network structure.

1 T. Martin, B. Ball, M.E.J. Newman, *Structural inference for uncertain networks*, Phys. Rev. E 93(1), 012306 (2016).

Topical homophily in online social systems

16

<u>Felipe M. Cardoso</u>[1,2], Sandro Meloni[1,3], André Santanchè[2], Yamir Moreno[1,3,4]

[1] Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, Spain
[2] Institute of Computing, State University of Campinas, Brazil
[3] Department of Theoretical Physics, University of Zaragoza, Spain
[4] Complex Networks and Systems Lagrange Lab, Institute for Scientific Interchange, Turin, Italy

In this work, we study if Twitter users tend to be in contact with people interested in similar topics, i.e., topical homophily. To do so, we employ a novel approach to model users through information topics extracted from Twitter messages. Our results show that, on average, users are connected with other users similar to them and stronger relationships are due to a higher topical similarity. Furthermore, we show that topical similarity provides interesting information for inferring users' relationships. We believe that this research, besides providing a way to assess the topical similarity of users, deepens the evidence of homophily among individuals, contributing to a better understanding of how complex social systems are structured.

Network Analysis of Commuting Flows in Poland

Mariola Chrzanowska[1]

[1]Department of Econometrics and Statistics, Warsaw University of Life Sciences, Poland

Commuting concerns peoples spatial behaviour resulting from the geographic separation of home and workplace. It is connected with their willingness to seek economic opportunities outside their place of residence [1]. The present paper offers a network analysis of the spatial commuting network in Poland. The main goal of this paper is to identify and analyse relations between communities (municipalities) from the perspective of labour market, especially commuting. The statistics of intercommunal data matrix concerning commuting scale will be used for the analyses. Data on the number of commuters come from the Central Statistical Office of Poland and cover the years 2006 and 2011. Spatial interaction model will be applied as a research method.

1 J. Rouwendal, P. Nijkamp, *Living in Two Worlds: A Review of Home-to-Work Decisions*, *Growth and Change*, Volume 35, Issue 3, p. 287. (2004).

From multimodal images to connectivity of brain networks in both healthy and pathological conditions

Jesus M Cortes[1,2,3,4,5]

[1] Ikerbasque: The Basque Foundation for Science. Bilbao, Spain
[2] Biocruces Health Research Institute. Barakaldo. Spain
[3] Department of Cell Biology and Histology. University of the Basque Country. Leioa, Spain
[4] This work has been done in the Computational Neuroimaging Lab at the Biocruces Health Research Institute in collaboration to several other members, including Asier Erramuzpe, Ibai Diez, Iaki Escudero, Beatriz Mateos, Sebastiano Stramaglia and Paolo Bonifazi. Other coauthors can be found in the cited references.
[5] JMC acknowledges financial support from Ikerbasque (The Basque Foundation for Science) and from the Ministerio Economia, Industria y Competitividad (Spain) and FEDER, grant DPI2016-79874-R.

To study brain functioning and disfunctioning the Computational Neuroimaging Laboratory in Biocruces (Bilbao) is focused on creating and applying brain connectivity methods obtained from magnetic resonance imaging. In particular, we focus on brain networks at the macroscale, thus affecting to the brain as a whole and not to specific brain areas. While diffusion images provide information about structural connectivity, functional magnetic resonance imaging shed light on resting state functional connectivity. In this talk I will focus on recent research focusing on the relationship of both functional and structural networks, and how these methods can be applied to study disconnectivity after a brain insult, such as traumatic brain injury.

The emergence of core-cohesive peripheries blockmodel type

Marjan Cugmas, Aleš Žiberna, Anuška Ferligoj

Faculty of Social Sciences, University of Ljubljana

The study proposes the core-cohesive peripheries blockmodel type, consisting of one highly popular group of units and two or more cohesive groups of units. Based on the previous studies, regarding friendship networks of preschool childrens, the popularity and transitivity mechanisms are assumed to be related with the emergence of the core-cohesive peripheries blockmodel type. Therefore, two main hypotheses are tested: (i.) the core-cohesive peripheries blockmodel can emerges from random network by the influence of popularity and transitivity mechanisms, and (ii.) the core-cohesive peripheries blockmodel can emerges from random network through core-periphery, since the strength of the popularity and transitivity is assumed not to be constant in time. The hypotheses are tested using the Monte Carlo simulations. The relative criterion function is used to compare the fits of different blockmodel types on empirically generated networks.

From Networks to Communities: a Robustness Measure

Annamaria Carissimo[1], Luisa Cutillo[2,3], Italia Defeis[4]

[1] Bioinformatics Core, TIGEM, Pozzuoli, Italy
[2] DCS, University of Sheffield [3] DISAQ, University of Naples "Parthenope"
[4] Istituto per le Applicazioni del Calcolo "Mauro Picone", Naples, Italy

We address the important question of the statistical validation of Network community detection. In this paper we present a methodology able to clearly detect if the community structure found by some algorithms is statistically significant or is a result of chance, merely due to edge positions in the network. Given a community detection method and a network of interest, our proposal examines the stability of the partition recovered against random perturbations of the original graph structure. To address this issue, we specify a perturbation strategy and a null model to build a set of procedures based on a special measure of clustering distance, namely Variation of Information, using tools set up for functional data analysis. The procedures determine whether the obtained clustering departs significantly from the null model. We show the results obtained with the proposed technique on simulated and real datasets.

Network analysis of commuting flows for delimitation of functional urban areas

Nina Drejerska[1], Mariola Chrzanowska[2]

[1]Department of European Policy and Marketing
[2]Department of Econometrics and Statistics,
Warsaw University of Life Sciences, Poland

Usually work opportunities can be found in the urban areas and this direction of flows is often a subject of research of urban studies or research focusing on city centers [1]. However, there is also such a phenomenon as reverse commuting directed from city centers to suburban areas. These two directions of interactions can be subjects of a network analysis. The main aim of the paper is to identify commuting patterns from suburbs to a city center and reversed ones. Data on the number of commuters come from the Central Statistical Office of Poland and cover the year 2011

1 N. Drejerska, M. Chrzanowska, *Commuting in the Warsaw suburban area from a spatial perspective - an example of empirical research*, Folia Oeconomica 2014, Vol. 6, no 309, pp. 87-96. (2014).

A novel approach to network anomaly detection

Mihai Cucuringu[1,2], Charlotte Deane[1,2], Andrew Elliott[1,3], Gesine Reinert[1,2].

[1] The Alan Turing Institute, UK.
[2] Department of Statistics, University of Oxford.
[3] CABDyN Complexity Centre, University of Oxford.

Presentation of work in progress on anomaly detection in networks (including time dependent networks) using spectral methods and manifold learning, which we test on both simulated and real world data. Our methods have applications in several domains including transaction networks, financial networks and cyber security.

How population clustering within a country affects infectious disease spread dynamics

Solveig Engebretsen[1], Birgitte F. de Blasio[1], Arnoldo Frigessi[1], Kenth Engø-Monsen[2]

[1] Department of Biostatistics, University of Oslo, Norway
[2] Telenor Research, Telenor Group, Fornebu, Norway

In many real countries, there is clustering in population sizes in that cities with large population sizes are often surrounded by other cities of large population sizes and vice versa. The effect of this clustering of population on infectious disease dynamics is studied by simulating infectious disease spread in different fictional countries, where the spatial correlation of population size is controlled. The disease dynamics model is a metapopulation model which can be described as a network where every node represents a location, and every edge represents people who travel between the locations and thus can spread the disease to new locations. In every location there is a separate set of stochastic differential equations governing the local disease dynamics, but the processes are coupled through travellers. The disease dynamics for the different countries are compared by considering various characteristics such as for instance spreading probabilities, arrival dates, final sizes and durations.

Gromov-Wasserstein distance of complex networks

Juan Fernández Gracia[1], Raúl Vicente [2], Víctor M. Eguíluz[1]

[1] IFISC (UIB-CSIC), Palma de Mallorca, Spain
[2] Institute of Computer Science, University of Tartu, Tartu, Estonia

One is often interested in comparing sets of networks coming from similar settings or describing similar interactions. Such a measure can provide a way to inspect the relations that the networks themselves have between each other. However, it is unclear which measure is more appropriate for comparing two or more networks. For example, it is difficult to compare two networks when the number of nodes is different or more generally, when we lack a correspondence between the nodes of one network and the nodes of the other. Here we investigate Gromov-Wasserstein distance (a metric distance between metric spaces) applied to complex networks. The Gromov-Wasserstein distance aims to solve a certain graph isomorphism problem by finding an optimal mapping between two sets of nodes such that their respective distances are best preserved. The deviation from a perfect isomorphism is related to the distance between the two objects. In particular we investigate its properties on network models and real biological networks.

Non-homogeneous dynamic Bayesian networks with partially edge-wise coupled interaction parameters

MAHDI SHAFIEE KAMALABAD[1] AND <u>MARCO GRZEGORCZYK</u>[1]

[1]Johann Bernoulli Institute, Groningen University, Netherlands

Non-homogeneous dynamic Bayesian networks (NH-DBNs), where a multiple changepoint process is used to divide the temporal data into disjunct segments, have become a popular tool for modelling time-varying regulatory networks. In the standard uncoupled NH-DBN the segment-specific parameters have to be learned separately for each individual segment, even if parameters stay similar. Recently NH-DBNs with coupled network parameters have been proposed. Parameter coupling leads to significantly improved network reconstruction accuracies when the segment-specific parameters are similar. However, recently we have found that coupling can become very counter-productive if there are parameters which substantially change from segment to segment. We therefore developed a new NH-DBN model which combines the features of the uncoupled and the coupled NH-DBN to improve the performance of both. Instead of enforcing all edges to be coupled, our model operates edge-wise and infers for each individual edge whether the associated parameters should be coupled or stay uncoupled.

When shape matters: Brain networks studied under a persistent homology view

ESTHER IBÁÑEZ-MARCELO[1], ANGKOON PHINYOMARK[1]
AND GIOVANNI PETRI[1]

[1] ISI Foundation, Torino, Italy

Over the last decade, a number of techniques rooted in algebraic topology have been proposed as novel tools for data analysis and pattern recognition [1,2]. These tools, referred to as topological data analysis (TDA), abandon standard measures between data points (or nodes) as the fundamental building block, and focus on extracting and understanding the shape of data at the mesoscopic scale. We leverage them to provide evidence of a large-scale topological preservation in brain functional topology mediated by a local reorganization of the homological localization across various physiological and altered brain states: sleep, LSD, psilocybin and their control groups.

1 H. Edelsbrunner, J. Harer. *Persistent Homology a Survey*, Contemporary Mathematics, 453:257282, (2008).

2 A. Zomorodian, G. Carlsson. *Computing Persistent Homology*, Discrete & Computational Geometry, **33**(2):249–274, (2005).

Quantifying and comparing success in artistic careers

Milan Janosov[1], Roberta Sinatra[2]

[1] Center for Network Science, Central European University, Hungary

When will an artist create her most successful piece? Does career success evolve differently in distinct fields? It has recently been shown by examining a big dataset of scientific careers that exceptional success occurs uniformly randomly over a career (Sinatra et al. 2016). This finding has been named random-impact-rule, and has been used as the base of a quantitative model, the Q-model, which provides long-term predictions of scientific impact. Here we build on this methodology to study careers in artistic fields. To this end, we created a large career database of tens of thousands of individuals from the book, film, and music industries. After confirming the random-impact-rule in all studied fields, we plan on using the Q-model to untangle the probabilistic and deterministic components of success. With this deconvolution we want to probe the presence of intrinsic differences between the studied fields, by quantitatively comparing the probabilistic and deterministic components of success in each of them. Finally, we aim to extend the Q-model to capture the effect of collaboration on success.

What do infection incidences tell us about connectivity of a regional healthcare network?

Andzej Jarynowski[1] and Vitaly Belik[2]

[1] Interdisciplinary Research Institute, Wroclaw, Poland
[2] System Modeling Group, Institute for Veterinary Epidemiology and Biostatistics, Freie Universität Berlin, Germany

We investigate the nosocomial infections registry of Malopolska Region (Lesser Poland), which contains more than 10.000 infection events for various alert pathogens during hospitalization in 2015 from around 60 hospitals. Database is used to build a spatial network, where vertices are hospital locations and edges represent similarity of epidemiological situation. Clustering as well as spatial correlations using agglomerative hierarchical structure detection techniques were performed. Our analysis indicates that one border county 'Oswiecim' does not belong to the rest of regional network. Moreover, the similarity in the incidence of viral pathogens seems to decay with distance much stronger than in bacterial pathogens. We also extend our approach to the dynamic case by a retrospective analysis of a sample of 25 hospitals in 2013-2015. Our analysis and modeling of a regional healthcare network facilitates decision making in epidemic control from a system perspective.

On the Graphical Markov Models with an Application

CEM KADILAR[1], GAMZE KADILAR[2]

[1,2] Department of Statistics, Hacettepe University, Turkey

Markov random fields or undirected graphical models are widely used for modeling, visualization, inference, and exploratory analysis of multivariate data with wide-ranging applications. Graphical models are models based on graphs in which nodes represent random variables, and the edges represent conditional independence assumptions. Hence they provide a compact representation of joint probability distributions. Graphical Markov models are multivariate statistical models which are currently under vigorous development and which combine two simple but most powerful notions, generating processes in single and joint response variables and conditional independences captured by graphs. The development of graphical Markov started with work by Wermuth (1976, 1980). The study of these models is an active research area, with many questions still open. In this study, how graphical Markov models emerged in the last 40 years are described. Then, interpretations are illustrated with an application. Furthermore, some of the more recent, important results for sequences of regressions are summarized.

An Analysis of Air Quality Data with Markov Chains and Time Series

GAMZE KADLAR[1], CEM KADLAR[2]

[1] Department of Statistics, Hacettepe University, Turkey
[2] Department of Statistics, Hacettepe University, Turkey

Sulphur dioxide ($SO2$) is a major air pollutant caused by the dominant usage of diesel, petrol and fuels by vehicles and industries. in this study, the level of $SO2$ is analyzed for several regions of Turkey based on the database monitored at air quality monitoring station of Turkey. First, a time series analysis is used to forecast the level of $SO2$ air quality parameter. Then, air quality index is used to construct a Markov Chain Analysis to predict the level of air quality parameter. The results indicate that the seasonal ARIMA model provides reliable and satisfactory predictions for the air quality parameters and expected to be an alternative tool for practical assessment and justification.

Zealots in the mean-field noisy voter model

Nagi Khalil[1], Maxi San Miguel[1], Raul Toral[1]

[1] IFISC (CSIC-UIB), Instituto de Física Interdisciplinar y Sistemas Complejos, Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain

The influence of zealots on the noisy voter model is studied theoretically and numerically at the mean-field level. The noisy voter model is a modification of the voter model that includes a second mechanism for transitions between states: apart from the original herding processes, voters may change their states because of an intrinsic, noisy in origin source. By increasing the importance of the noise with respect to the herding, the system exhibits a finite-size phase transition from a quasi-consensus state, where most of the voters share the same opinion, to a one with coexistence. Upon introducing some zealots, or voters with fixed opinion, the latter scenario may change significantly. We unveil the new situations by carrying out a systematic numerical and analytical study of a fully connected network for voters, but allowing different voters to be directly influenced by different zealots.

1 Nagi Khalil, Maxi San Miguel, and Raul Toral, ArXiv:1707.04087

Random order statistics probability models for networks

Danai Koutra[1], <u>Markos V. Koutras</u>[2],

[1] Department of Computer Science & Engineering, University of Michigan, Ann Arbor, USA
[2] Department of Statistics & Insurance Science, University of Piraeus, Greece

Let N be a non-negative integer valued random variable and $X_1, X_2, \ldots$ a sequence of independent and identically distributed random variables, independent of N. The analysis of many problems associated with the study of social networks as well as the exploration of several problems of applied nature that can be modelled by the aid of graphs, require the study of the distribution of appropriate statistics defined over the random sample $X_1, X_2, , X_N$ whose length N is a random variable.

In the present work we discuss the exact distribution of order statistics of random samples of random lenght and elucidate how the resuls can be exploited in modelling networks/graphs and extracting useful information both for their structure and possible similarities/dissimilarities between them.

International Arms Trade: A Dynamic Separable Network Model with Heterogeneity Components

Michael Lebacher[1], Paul W. Thurner[2], Göran Kauermann[1]

[1] Department of Statistics, Ludwig-Maximilians-University Munich
[2]Geschwister Scholl Institute of Political Science, Ludwig-Maximilians-University Munich

We investigate data from the Stockholm International Peace Research Institute (SIPRI) on international trade of major conventional weapons from 1950 to 2016, using a network model. The model is based on the separable temporal exponential random graph model (StERGM), but extends this model in two aspects. In order to account for dynamics, we allow for time-varying covariate effects. The considerable actor- and time-based heterogeneity is modelled by incorporating smooth time-varying random effects. In a second step, the random effects are subjected to a functional principal component analysis. Our main findings are that arms trading is driven by strong network effects, notably reciprocity and triadic closure, but also by political and economic characteristics. A careful analysis of the country-specific random effects identifies countries that increased or decreased their relative importance in the arms network during the observation period.

A Network Epidemic Model for Online Community Commissioning Data

Clement Lee[1,2], Andrew Garbett[1], Darren J. Wilkinson[2]

[1] Open Lab, Newcastle University, UK
[2] School of Mathematics and Statistics, Newcastle University, UK

A statistical model assuming a preferential attachment network, which is generated by adding nodes sequentially according to a few simple rules, usually describes real-life networks better than a model assuming, for example, a Bernoulli random graph, in which any two nodes have the same probability of being connected. Therefore, we propose a network epidemic model by combining a stochastic epidemic model and a preferential attachment model. A simulation study based on the subsequent Markov Chain Monte Carlo algorithm reveals an identifiability issue with the model parameters. Finally, the network epidemic model is applied to a set of online commissioning data.

ERGMs in neuroimaging - developments and challenges

Brieuc Lehmann

Moment-based parameter estimation in binomial random intersection graph models

Joona Karjalainen, <u>Lasse Leskelä</u>

Department of Mathematics and Systems Analysis
Aalto University, Finland

Binomial random intersection graphs are parsimonious statistical models of large and sparse networks, with one parameter for the average degree and another for transitivity, the tendency of neighbours of a node to be connected. We will discuss the estimation of these parameters from a single observed instance of the graph, using moment estimators based on observed degrees and frequencies of 2-stars and triangles. The observed data set is assumed to be a subgraph induced by a set of $n_0 \gg n^{2/3}$ nodes sampled from the full set of n nodes. We prove the consistency of the proposed estimators by showing that the relative estimation error is small with high probability. The talk is based on [1].

1 J. Karjalainen, L. Leskelä, *Moment-based parameter estimation in binomial random intersection graph models*, arXiv:1704.04278.

Graphoid properties and independence with Credal Networks

SABINA MARCHETTI[1],

[1] Department of Statistics, Sapienza University, Rome

Graphical models are popular statistical tools to model uncertainty, combining probability and graph theory. Networks serve for representing local dependencies among a (possibly large) set of random variables, making the inference process efficient. Credal Networks generalize directed graphical models to the setting of imprecise beliefs, where probability values are allowed to vary within intervals/ranges of probabilities. There, different forms of independence ought to be considered, together with their graphical representation and properties. We provide an up-to-date literature review on the subject.

1 F.G. Cozman, *Credal Networks*, Artificial intelligence 120.2 (2000): 199-233.

2 P. Walley, *Statistical reasoning with imprecise probabilities*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis (1991).

The Multilayer Networks for Language

SANDA MARTINČIĆ-IPŠIĆ[1], ANA MEŠTROVIĆ[1]

[1] Department of Informatics, University of Rijeka, Croatia

We introduce the multilayer networks for language–a unified framework for modeling linguistic subsystems and their structural properties and exploring their mutual interactions [1]. Language can be represented as complex networks: vertices are linguistic units and links their relations. The language multilayer network is defined by three aspects: network construction principle, linguistic subsystem and language of interest. More precisely, we construct a word-level (syntax, shuffled and co-occurrence) and a subword-level (syllables and graphemes) network layers, from five variations of English and Croatian texts. The results show that there are differences in the structures of lang. subsystems, which are hidden during the exploration of an isolated layer. The word-level layers share structural properties regardless of the language, while subword-level expresses language dependent properties.

1 S. Martinčić-Ipšić, D. Margan, A. Meštrović, *Multilayer Network of Language: a Unified Framework for Structural Analysis of Linguistic Subsystems.* Physica A, 457, pp. 117-128,(2016).

A semiparametric extension of the stochastic block model for longitudinal networks

C. Matias, T. Rebafka, F. Villers

To model recurrent interaction events in continuous time, an extension of the stochastic block model is proposed where every individual belongs to a latent group and interactions between two individuals follow a conditional inhomogeneous Poisson process with intensity driven by the individuals' latent groups. The model is shown to be identifiable and its estimation is based on a semiparametric variational expectation-maximization algorithm. Two versions of the method are developed, using either a nonparametric histogram approach (with an adaptive choice of the partition size) or kernel intensity estimators. The number of latent groups can be selected by an integrated classification likelihood criterion. Finally, we demonstrate the performance of our procedure on synthetic experiments, analyse two datasets to illustrate the utility of our approach and comment on competing methods.

A stochastic heterogeneous mean-field approximation of agent-based models

Jérôme Michaud[1]

[1] Department of Physics and Astronomy, Uppsala University, Sweden

In this talk I discuss a procedure to obtain a coarse-grained approximation of the dynamics of agent-based models (ABMs) defined on a complex network [1]. It is an heterogeneous mean-field approximation, since the dynamics is approximated by averaging the behavior of groups of agents (the grouping procedure controls the heterogeneity of the approximation). Our approximation captures the stochastic aspects of the dynamics, which are usually neglected by mean-field approximations. The approximation is validated by numerical simulations on the utterance selection model [2], which is an ABM for language change in a population, and shows excellent agreement.

1 J. Michaud *Continuous time limits of the utterance selection model*, Phys. Rev. E 95, 022308 (2017).

2 G. J. Baxter, R. A. Blythe, W. Croft, and A. J. McKane *Utterance selection model of language change*, Phys. Rev. E 73, 046118 (2006).

Fragmentation transitions in a coevolving nonlinear voter model

Byungjoon Min[1], Maxi San Miguel[1]

[1] IFISC, Instituto de Física Interdisciplinar y Sistemas Complejos (CSIC-UIB), Campus Universitat Illes Balears, E-07122 Palma, Spain

We study a coevolving nonlinear voter model describing the coupled evolution of the states of the nodes and the network topology. Nonlinearity of the interaction is measured by a parameter q. The network topology changes by rewiring links at a rate p. By analytical and numerical analysis we obtain a phase diagram in p, q parameter space with three different phases: Dynamically active coexistence phase in a single component network, absorbing consensus phase in a single component network, and absorbing phase in a fragmented network. For finite systems the active phase has a lifetime that grows exponentially with system size, at variance with the similar phase for the linear voter model that has a lifetime proportional to system size. We find three transition lines corresponding to different transition mechanisms that meet at the point of the fragmentation transition of the linear voter model.

Bayesian Structure Learning of High-dimensional Graphical Models with Application to Brain Connectivity

Reza Mohammadi

Dept. of Methodology and Statistics, Tilburg University, Netherlands

We consider the problem of Bayesian structure learning in high-dimensional graphical models, motivated by brain connectivity applications. In graphical models, Bayesian frameworks provide a straightforward tool, explicitly incorporating underlying graph uncertainty. In principle, the Bayesian approaches are based on averaging the posterior distributions of the quantity of interest, weighted by their posterior graph probabilities. However, Bayesian inference has not been used in practice for high-dimensional graphical models, because computing the posterior graph probabilities is hard and the number of possible graph models is very large. We discuss the computational problems related to Bayesian structure learning and we offer several solutions to cope the high-dimensionality problems. We apply our method to high-dimensional fMRI data from brain connectivity studies to show its empirical usefulness. In addition, we have implemented our method in the `R` package **BDgraph** which is available online.

FANTASI: Fast Network Analysis in Silicon

Andrey Mokhov[1], Alessandro de Gennaro[1],
Ghaith Tarawneh[1], Jonny Wray[2]

[1] Newcastle University, Newcastle upon Tyne, United Kingdom
[2] e-Therapeutics, Oxford, United Kingdom

We present a hardware accelerator for network analysis that achieves 1000x speed-up compared to a conventional 96-core computer cluster on the all-pairs shortest path problem. The key idea behind the approach is to embed networks in silicon (a field-programmable gate array or FPGA), where network analysis can be performed by propagating signals through the physical network and measuring signal propagation time. This eliminates the memory bottleneck and allows thousands of nodes to be processed in parallel [1]. The accelerator was developed for e-Therapeutics, a pioneer in computational drug discovery, to accelerate their analyses of protein-interaction networks.

1 A. Mokhov, A. de Gennaro, G. Tarawneh, J. Wray, G. Lukyanov, S. Mileiko, J. Scott, A. Yakovlev, A. Brown. *"Language and Hardware Acceleration Backend for Graph Processing"*, IEEE Forum on Specification & Design Languages (FDL), 2017.

Local inference by penalization method for biclustering model

Eduard Belitser[1], Nurzhan Nurushev[1]

[1] VU Amsterdam

We study the problem of inference on the unknown parameter in the *biclustering model* by using the penalization method. The underlying biclustering structure is that the high-dimensional parameter consists of a few blocks of equal coordinates. The main goal is the *uncertainty quantification*, but on the way we also solve the estimation problem as well. We pursue a novel *local* approach in that the procedure quality is characterized by a local quantity, *the oracle rate*, which is the best trade-off between the approximation error by a biclustering structure and the best performance for that approximating biclustering structure. The approach is also *robust* in that the additive errors are not assumed to be independent with some known distribution, but only satisfying certain mild conditions. We introduce the *excessive bias restriction* under which we establish the local (oracle) confidence optimality of the proposed confidence ball. Adaptive minimax results for the *graphon* estimation and *stochastic block model* follow from our local results with implications for network modeling.

From Ego Networks to Dunbar's Numbers and Social Structure: an experiment

María Pereda[1], Ignacio Tamarit[1], Jose A. Cuesta[1,2,3], Robin Dunbar [4], Anxo Snchez [1,2,3]

[1] Universidad Carlos III de Madrid. Mathematics department. GISC
[2] Institute UC3M-BS for Financial Big Data, Universidad Carlos III de Madrid, Spain
[3] Institute for Biocomputation and Physics of Complex Systems (BIFI), Zaragoza, Spain
[4] Department of Experimental Psychology, University of Oxford, Oxford, UK

The average size of a human social network is about 150 individuals. This limit has been connected to our cognitive capacity and it is known as Dunbar's Number. Furthermore, personal networks are typically arranged into a set of hierarchically inclusive layers whose sizes follow a characteristic sequence referred as Dunbar's circles: 5, 15, 50, 150. Little is known about its creation and early evolution. So, we have designed a longitudinal experiment in a (closed) community of +300 teenage students, from 11 to 18 years old. Our goal is to capture the social network of each individual at various levels of relationship, over time; and to collect information about personal traits that could potentially explain the way relationships are established and sustained. Here we present results from the first wave of our experiment. We find the two first circles predicted by the theory, and show how measures of structural overlapping between networks can serve as a predictor of the type of relationship between any two pair of individuals. We also discuss the features of the community structure of the networks.

Classifier Precision Analysis for Sleep Apnea Detection with ECG Signals

André Pinho[1], Nuno Pombo[1], Nuno Garcia[1,2], Koumana Bousson[3]

[1] Instituto de Telecomunicações, Universidade da Beira Interior, Covilhã, Portugal
[2] Universidade Lusófona de Humanidades e Tecnologias, Covilhã, Portugal
[3] Departamento de Ciências Aeroespaciais, Universidade da Beira Interior, Covilhã

This is a study on the efficiency of implementing classifiers and the use of features for the detection of sleep apnea moments based on a Electrocardiogram (ECG) signal, detailing the comparison of the different classifiers when presented with different amount of features, individually ranked using Discriminant Relevance (DR). Features were extracted based on the Heart Rate Variability (HRV and ECG-Derived Respiration (EDR). Training and testing sets were obtained by randomly dividing the data (set of features) using k-fold cross validation (k=10). The best accuracy obtained was 82.12%, with a sensitivity and a specificity of 88.41% and 72.29%, respectively. These promising results may lead to complementary studies to improve classifiers in order to test it with end-users and for validation under different scenarios.

Predicting language diversity with complex networks

Tomasz Raducha[1,2], Tomasz Gubiec[1]

[1] Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Poland
[2] Instituto de Física Interdisciplinar y Sistemas Complejos (CSIC-UIB), Palma de Mallorca, Spain

Evolution and propagation of the world's languages is a complex phenomenon, driven, to a large extent, by social interactions. Multilingual society can be seen as a system of interacting agents, where the interaction leads to a modification of the language spoken by the individuals. Here we address the issue of the language diversity in societies of different sizes, and we show that local interactions are crucial to capture characteristics of the empirical data. We propose a model [1] that cancels the contradiction between previous models and the Solomon Islands case. Our results demonstrate the importance of the topology of the network, and the rewiring mechanism in the process of language change.

1 T. Raducha, T. Gubiec, arXiv:1704.0835 (2017)

Mixture model under overlapping clusters: an application to network data

Saverio Ranciati[1], Veronica Vinciotti[2], Ernst C. Wit[1,3]

[1] Department of Statistical Sciences, University of Bologna, Italy
[2] Department of Mathematics, Brunel University London, The UK
[3] J.B.I. for Mathematics and Computer Science, University of Groningen, The Netherlands

In this paper, we present a finite mixture model for the general case of (potentially) overlapping clusters. We present a Bayesian inference approach for the estimation of the parameters and an application to the clustering of actor-event binary network data, whereby terrorists (actors) are clustered according to their preferential attendance to a set of events of different nature.

Bayesian networks for data imputation in survival tree analysis

PAOLA M.V. RANCOITA[1], MARCO ZAFFALON[2], EMANUELE ZUCCA[3], FRANCESCO BERTONI[3], CASSIO P. DE CAMPOS[4]

[1] CUSSB, Vita-Salute San Raffaele University, Milan, Italy
[2] IDSIA, Manno, Switzerland
[3] Oncology Institute of Southern Switzerland, Bellinzona, Switzerland
[4] Queens University Belfast, Belfast, UK

The objective of many clinical studies is prognostic patient stratification and this can be obtained by survival tree analysis. Moreover, often, retrospective data are employed for this purpose but they may present many missing covariates, besides censored data.

To overcome this issue, we propose an approach based on Bayesian networks for modeling the dependencies among covariates and, once the conditional distributions are estimated, for imputing the missing values. The proposed method generally outperformed other existing methods for data imputation on simulated data. In addition, it allowed a better estimation of the prognostic patient stratification in the survival tree analysis on both the simulated and real data.

Coinfection in the space: cooperative disease spreading in geometric and contact networks

JORGE P. RODRÍGUEZ[1], FAKHTEH GHANBARNEJAD[2] , VÍCTOR M. EGUÍLUZ[1]

[1] Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), Palma de Mallorca, Spain
[2] Technische Universität Berlin, Germany

Cooperative interactions between diseases lead to abrupt outbreaks in networks with a low ratio between the number of short and long loops [1]. Spatial networks represent realistic topologies to estimate the risks of big outbreaks caused by this coinfection dynamics. We study the cooperative disease spreading processes in three networks: first, a static geometric network, modifying the interaction radius; secondly, a mobile geometric network where, as the velocity is increased, an abrupt transition appears; finally, we simulate this dynamics in empirical contact networks.

1 W. Cai, L. Chen, F. Ghanbarnejad, P. Grassberger, *Avalanche outbreaks emerging in cooperative contagions*, Nat. Phys. 11, 936 (2015).

Plasticity on Ecological Bipartite Networks

SOMAYE SHEYKHALI[1,2], JUAN FERNNDEZ GRACIA[2], VCTOR M. EGULUZ[2]

[1] Department of Physics, University of Zanjan, Zanjan, Iran
[2] Instituto de Fsica Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), Campus UIB
07122 Palma de Mallorca, Spain.

One of the fundamental questions in ecology is what are the structural and functional properties of many-species systems that allow species persistence. A high level of nestedness and modularity are common features of bipartite ecological networks (e.g. plant-pollinator). Loss or declining abundance of one species can lead to extinctions of other dependent species. For instance, specialist predators will go extinct following the loss of their only prey unless they switch to a new species. It has therefore been suggested that an ability of consumers to rewire to novel prey should mitigate the consequences of species loss by reducing the risk of cascading extinction. In this study, we model population dynamics to evaluate the response of ecological bipartite networks to environmental variability and species loss as external perturbations. We do so by characterizing the evolution of the ecological bipartite network structure, focusing on nestedness and modularity, as we remove nodes from it and compare the performance of various rewiring rules.

Network Analysis of scientific Co-operation between faculties at Warsaw University of Life Sciences

DOMINIK SIERADZKI[1], MSc; MARIOLA CHRZANOWSKA[1], PhD

[1]Department of Econometrics and Statistics, Warsaw University of Life Sciences, Poland

The purpose of the poster was to characterise faculties at Warsaw University of Life Sciences as a network-based the largest agricultural university in Poland and the evaluation of capacity of network analysis in studying scientific cooperation. Results of the study shows that faculties create a dense, closely interconnected network of cooperation. The network is dominated by a limited number of division, which are involved in large share of the projects and have the most expanded cooperation network. Network analysis allowed to identify the lacks of scientific cooperation between some faculties.

How to integrate gene enrichment analysis with information from gene interaction networks

Mirko Signorelli[1], Veronica Vinciotti[2], Ernst C. Wit[3]

[1] Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, The Netherlands
[2] Department of Mathematics, Brunel University London, UK
[3] Johann Bernoulli Institute, University of Groningen, The Netherlands

Enrichment analysis allows to characterize differentially expressed gene sets, associating known biological functions to them. However, traditional enrichment analysis tests simply consider the number of overlapping genes among gene sets, ignoring functional couplings between genes. A solution to this problem is the integration of enrichment analysis with information from gene interaction networks, known as network enrichment analysis. In this talk we will present NEAT [1], a statistical test that allows to assess relations between gene sets in gene interaction networks in an efficient way, and compare it to the existing alternative tests for network enrichment analysis.

1 M. Signorelli, V. Vinciotti, E. Wit. *NEAT: an efficient network enrichment analysis test*, BMC Bioinf., 17:352 (2016).

Three forms of convexity in graphs and networks

Tilen Marc[1], Lovro Šubelj[2]

[1] Institute of Mathematics, Physics and Mechanics, Ljubljana, Slovenia
[2] University of Ljubljana, Ljubljana, Slovenia

Convexity is a property of a part of mathematical object that includes all the shortest paths between its units. In the case of graphs or networks, a connected subgraph is convex if every shortest path between the nodes of the subgraph lies entirely within the subgraph. We define a convex network as such in which every connected subset of nodes induces a convex subgraph. We show that such convexity is an inherent property of many networks that is not present in random graphs [1]. In particular, spatial infrastructure networks and social collaboration graphs are globally convex due to their tree/clique-like structure. Core-periphery networks are regionally convex as they can be divided into a non-convex core surrounded by a convex periphery. Random graphs, however, are merely locally convex meaning that only subgraphs of size smaller than the average distance between the nodes are convex.

1 T. Marc, L. Šubelj, *Convexity in complex networks*, arXiv:1608.03402v3, pp. 27 (2016).

Emergence of social structure: a generalization of Dunbar's circles.

Ignacio Tamarit[1], Jose A. Cuesta[1,2,3], Robin Dunbar [4], Anxo Sánchez [1,2,3]

[1] Universidad Carlos III de Madrid. Mathematics Department. GISC
[2] Institute UC3M-BS for Financial Big Data, Universidad Carlos III de Madrid, Spain
[3] Institute for Biocomputation and Physics of Complex Systems (BIFI), Zaragoza, Spain
[4] Department of Experimental Psychology, University of Oxford, Oxford, UK

Humans have social networks with a characteristic size and layered structure. Typically, an individual deals with about 150 relationships organised into a set of hierarchical layers of increasing size with decreasing emotional intensity. Furthermore, these inclusive layers seem to exhibit a constant scaling ratio. While the precise size of the networks has been connected to our cognitive capacity, no mechanism explaining its structure has been proposed. Here we show that the existence of a limited cognitive capacity and heterogeneity in the relationships imply, without further assumptions, the existence of layers and the scaling between them. To that end, we develop a Bayesian model capable to fit empirical data remarkably well and predict the existence of a new, unexpected regime in the case of small communities. We also provide empirical evidence of the existence of such regime. Thus, our model explains the emergence of a social (macroscopic) structure based on individual (microscopic) features and predicts a novel social phenomenon that is later confirmed empirically.

PC algorithm for complex survey data via resampling

Daniela Marella[1], Paola Vicard[2]

[1] Department of Economics, University Roma Tre, Italy
[2] Department of Education Science, University Roma Tre, Italy

PC algorithm is one of the most known procedures for Bayesian network structural learning. PC algorithm uses conditional independence tests and it is based on the assumption of $i.i.d.$ observations, equivalent to simple random sampling assumption. It is well known that for $i.i.d.$ data the Pearson chi-squared test statistic is asymptotically distributed as a chi-squared random variable under the null hypothesis. For complex designs, the test procedure is not valid even asymptotically. In fact, the impact of complex designs on $i.i.d.$ based methods can be severe.

Here, a novel approach for inferring casual structure from complex survey data is investigated. The complexity of sampling design is accounted for in a design-based approach by including the sampling weights in the BN parameters estimates. A modified version of the PC algorithm, named PC complex is proposed. It is based on resampling techniques for finite population. PC complex is applied to BN structural learning and its robustness with respect to the departure from faithfulness assumption is studied.

Motion of substances in channels of networks

Nikolay K. Vitanov[1], Zlatinka I. Dimitrova[2], Roumen Borisov[1]

[1] Institute of Mechanics, Bulgarian Academy of Sciences, Sofia, Bulgaria
[2] Institute of Solid State Physics, Bulgarian Academy of Sciences, Sofia, Bulgaria

We model the motion of a substance in a channel of a network that consists of chain of nodes of the network and edges that connect the nodes. Some amount of the substance can leave the channel at a node and the rate of leaving can be different for the different nodes of the channel. The nodes close to the end of the channel may be more "attractive" for the substance in comparison to the nodes around the incoming node of the channel. We discuss channels containing infinite or finite number of nodes. The distribution of the substance along the nodes of the channel is a very long tail distribution that contains as particular case the Waring distribution (for channel with infinite number of nodes) or the truncated Waring distribution (for channel with finite number of nodes)[1].

1 N. K. Vitanov, K. N. Vitanov, *Box model of migration channels*, Math. Soc. Sciences. 80, 108 (2016).

Network inference in genomics

Ernst C. Wit[1]

[1] Johann Bernoulli Institute of Mathematics and Computer Science, University of Groningen, The Netherlands

The whole concept of network inference in genomics has multiple meanings and interpretations. It can refer to "causal" or "topological" considerations, i.e., learning about functional relationships in the genomic system or to considerations about the structure of the overall genomic network. Moreover, the genomic network does not really exists and can refer to gene regulatory networks, cell signaling networks, metabolic networks etc.

In this talk I aim to clarify the underlying genomics in order to motivate a hierarchy of four network inference strategies, starting at the single cell level and finishing at global structural network inference. It will involve stochastic and ordinary differential equation models, causal inference and graphical modelling.