Coarse Graining of Complex Systems

a satellite workshop of CCS 2016
(Amsterdam)

BOOK OF ABSTRACTS

Organizers:

Mauro Faccin,
Leto Peel,
Michael Schaub,
Renaud Lambiotte

Contents

1 Introduction

A fundamental issue in the field of complex systems is to identify the most important interactions or processes within a system so that we may better understand them. Coarse- graining methods attempt to abstract away unnecessary details, leaving behind a simplified interpretable representation of the system. Coarse-graining approaches are well-established in many different areas and have found applications for many different problems: e.g., clustering of high-dimensional data, reducing the complexity of time-series, or finding communities in networks. However, real-world complex systems present new challenges in this direction. Importantly, such systems often comprise multiple spatial and temporal scales that need to be integrated to establish a comprehensive system description. Furthermore, complex systems are often also characterised in terms of multiple types of interaction and available information. For instance, the spread of a disease through a population depends on the network of physical contacts, the time-scales at which they occur and the characteristics of the individuals.

The objective of this satellite session is to bring together theoretical and applied researchers to discuss challenges and new approaches towards coarse-graining of such complex systems. Areas of interest include but are not limited to:

- Dynamical systems on networks, like disease spreading, transport, etc.

- Coarse-graining of rich, multi-relational data (web documents, social networks)

- Time series and temporal (network) dynamics, e.g., in the brain

- Dynamics on and of socio-economic networks (finance, trade, etc.)

1 INTRODUCTION

1.1 Program: Wednesday 21 September 2016

8:30 — Keynote Lecture (from the conference program)

9:00 — Ignites

9:35 — *Maps of sparse Markov chains efficiently reveal community structure in network flows with memory*
L. Bohlin, C. Persson, D. Edler, M. Rosvall

10:10 — *Community Detection for Correlation Matrices*
D. Garlaschelli

10:30 — Coffee Break

11:00 — *Link transmission centrality in complex networks*
M. Karsai

11:35 — *Coarse-graining of Multiplex Networks*
E. Cozzo, R. Sanchez-Garcia, Y. Moreno

11:55 — *Turing patterns in time varying networks*
T. Carletti, D. Fanelli

12:15 — No-Coffee Break

12:20 — *Modularity and the spread of perturbations in complex dynamical systems*
A. Kolchinsky, A. Gates, L. Rocha

12:40 — *Local coarse-graining of network dynamics*
J. Delvenne, Y. Yu, S. Yaliraki, M. Barahona

13:00 — Lunch

14:00 — Keynote Lecture

14:50 — *Coarse graining and data aggregation techniques in location-based services*
A. Noulas

15:25 — *Impact of Network Structure on Knowledge Diffusion between Groups of OSS Developers*
S. Behfar, E. Turkina, T. Burger-Helmchen

15:45 — *Automatic identification of relevant concepts in scientific publications*
A. Martini, A. Cardillo, P. De Los Rios

16:05 — Coffee Break

16:35 — *Tensorial Stochastic Block Models for layered data*
M. Sales-Pardo

17:10 — *The dynamics of community sentiments on Twitter*
D. Vukadinovic Greetham, B. Charlton, C. Singleton

17:30 — *Probabilistic and flux-based analysis of metabolic graphs*
M. Beguerisse Diaz, M. Barahona, G. Bosque Chacón, D. Oyarzún, J. Picó

2 Invited Speakers

2.1 Maps of sparse Markov chains efficiently reveal community structure in network flows with memory

Ludvig Bohlin
Christian Persson
Daniel Edler
Martin Rosvall
Umeå University

To better understand the flows of ideas or information through social and biological systems, researchers develop maps that reveal important patterns in network flows. In practice, network flow models have implied memoryless first-order Markov chains, but recently researchers have introduced higher-order Markov chain models with memory to capture patterns in multi-step pathways. Higher-order models are particularly important for effectively revealing actual, overlapping community structure, but higher-order Markov chain models suffer from the curse of dimensionality: their vast parameter spaces require exponentially increasing data to avoid overfitting and therefore make mapping inefficient already for moderate-sized systems. To overcome this problem, we introduce an efficient cross-validated mapping approach based on network flows modeled by sparse Markov chains. To illustrate our approach, we present a map of citation flows in science with research fields that overlap in multidisciplinary journals. Compared with currently used categories in science of science studies, the research fields form better units of analysis because the map more effectively captures how ideas flow through science.

2.2 Link transmission centrality in complex networks

Marton Karsai
ENS de Lyon

The identification and control of ties, which are transmitting information in complex networks can provide effective ways to hinder and postpone outbreaks of spreading phenomena. In this work we introduce a new method by using diffusion processes to define transmission centrality for links, which quantifies their importance in disseminating information in the network structure. We demonstrate that this measure can precisely identify weak ties in the network structure and furthermore estimates the betweenness centrality of each link in a computationally efficient way. By testing various link removal strategies in three human communication networks we show that the combined measure of transmission centrality and link overlap provides the best way to identify links keeping the global structure connected. We also explored their role in control of spreading processes through SIR model simulations. This method allows us to select the optimal but minimal set of links to gain maximum control on spreading processes.

2.3 Local coarse-graining of network dynamics

Jean-Charles Delvenne
Université Catholique de Louvain
YW Yu
MIT
S. Yaliraki
Imperial College of London
M. Barahona
Imperial College of London

Usual coarse-graining techniques require a full-knowledge of the network and its dynamics. This assumption is increasingly irrealistic as we consider larger and larger networks, due to both incomplete data and lack computational resources. We propose a theory of local coarse graining where a set of nodes is reduced to a single node without knowledge of the rest of the network, and with limited impact on the resulting dynamics. This allows us to identify interesting structures in a network, including, but not limited to, local communities. The methodology is illustrated on social, biological and power networks. Joint work with YW Yu, S Yaliraki, M Barahona

2.4 Coarse graining and data aggregation techniques in location-based services

Anastasios Noulas
Data Science Institute at Lancaster University

Location-based services have become a popular subject of research over the past decade thanks to their significance as a novel source of geo-referenced data that provide solutions for a variety of research problems in a host of disciplines. Despite their richness in terms of the multiple data layers that become available, these datasets are often sparse and are characterised by skewed distributions which make the application of classical statistical frameworks and machine learning algorithms in this context a challenge. In this talk, we will review a wider range of data aggregation and coarse graining techniques that enable useful characterisations of complex location-based systems and find applicability in many real world applications.

2.5 Tensorial Stochastic Block Models for layered data

Marta Sales-Pardo
Universitat Rovira i Virgili

In this talk I will discuss the problem of developing predictive models for layered data. Time-resolved networks are a typical example of layered data, since each time window results in a specific pattern of connections. I will present stochastic tensorial block models as a valid approach to predict missing information in network data with different layers of information. I will discuss results for two cases: a temporally resolved e-mail communication network and a drug-drug interaction network in different cell lines.

3 Contributed Talks

3.1 Community Detection for Correlation Matrices

Diego Garlaschelli
University of Leiden

A challenging problem in the study of complex systems is that of resolving, without prior information, the emergent, mesoscopic organization determined by groups of units whose dynamical activity is more strongly correlated internally than with the rest of the system. The existing techniques to filter correlations are not explicitly oriented towards identifying such modules and can suffer from an unavoidable information loss. A promising alternative is that of employing community detection techniques developed in network theory. Unfortunately, this approach has focused predominantly on replacing network data with correlation matrices, a procedure that we show to be intrinsically biased because of its inconsistency with the null hypotheses underlying the existing algorithms. Here, we introduce, via a consistent redefinition of null models based on random matrix theory, the appropriate correlation-based counterparts of the most popular community detection techniques. Our methods can filter out both unit-specific noise and system-wide dependencies, and the resulting communities are internally correlated and mutually anticorrelated. We also implement multiresolution and multifrequency approaches revealing hierarchically nested subcommunities with "hard" cores and "soft" peripheries. We apply our techniques to financial and brain time series and identify robust patterns providing nontrivial information about these systems.

Ref: Mel MacMahon and Diego Garlaschelli, Phys. Rev. X 5, 021006 (2015).

3.2 Coarse-graining of Multiplex Networks

Emanuele Cozzo
BiFi — Universidad Zaragoza
Ruben Sanchez-Garcia
University of Southampton
Yamir Moreno
BiFi — Universidad Zaragoza

We understand multiplex networks as a non-linear superposition of complex networks, where components, being them social actors, genes, devices, or physical locations, interacts through a variety of different relationships and communication channels, which we conceptualize as different layers of the multiplex network. Because of the structure of a multiplex network, it is natural to try to aggregate the interaction pattern of each layer in a single network somehow. An operation that we call dimensionality reduction, whereas the result of such operation leads to what we call an aggregate network. Several candidates for the aggregate network have been proposed in the literature. We claim that the natural definition of an aggregate network is given by the notion of quotient network. In addiction, in the quotient network framework, we are able to introduce in a symmetric way another reduced network, the network of layers, that encodes the connectivity pattern between layers. We give the relations between the Laplacian and the adjacency spectra of the original multiplex network and that of its coarse-grained representations, and we show that a number of the structural and dynamical properties of the original multiplex are inherited by the coarse-grained representations. Finally, we show the relation between some walk-based structural metrics (e.g. the clustering coefficient) calculated on the original multiplex and the same metric calculated on the aggregate network.

3.3 Turing patterns in time varying networks

Timoteo Carletti
Namur Center for Complex Systems, University of Namur
Duccio Fanelli
University of Florence

Patterns are widespread in nature and appear in large plethora of different conformations.From chemistry to biology,passing through physics,ecology,social systems or engineered ones, just to name few.Beautiful spatially extended motifs are found which spontaneously emerge from an ensemble made of interacting microscopic actors. The proto-typical approach to patterns formation in reaction-diffusion processes dates back to Turing's seminal paper on morphogenesis[1].Under specific conditions,diffusion drives an instability by perturbing a homogeneous stable fixed point, via an activator-inhibitor mechanism.As the perturbation grows,nonlinear reactions balance the diffusion terms,yielding the asymptotic, spatially inhomogeneous, steady state. Reaction diffusion models are often defined on a regular lattice, either continuous or discrete.In many cases of interest, it is however more natural to schematize the system as a complex network.Building on the pioneering work of Othmer and Scriven[2], Nakao and Mikhailov developed[3] the theory of Turing patterns on random static undirected network, highlighting the peculiarities that stem from the embedding graph structure.The dispersion relation, from which the instability conditions ultimately descend, is related to the spectrum of the discrete Laplacian associated to the embedding network.Laplacian eigenvalues determine hence the spatial characteristics of the emerging patterns. Evidence suggests that in many applications networks evolve in time,new nodes can enter or other can leave the system,new links can emerge or disappear[4–6].Linear processes such as diffusion have been studied on temporal networks[7, 8].The goal of this paper is to propose a first study of a generic non-linear process running on top of a complex temporal network and determine the conditions for the onset of Turing patterns.We will provide examples where patterns emerge in the temporal network while they cannot develop in each time-snapshot network, namely for any fixed time.

[1]A.M.Turing, PTR.Soc.London, Ser.B 237,37(1952)
[2]H.G.Othmer, L.E.J.Scriven, Theor.Biol. 32,507(1971);43,83(1974)
[3]H.Nakao,A.S.Mikhailov, NatPhys 6,544(2010)
[4]V. Kostakos,PhysicaA 388,1007(2009)
[5]P. Holme,J.Saramäki,PhysRep 519,97(2012)
[6]P.Holme,EPJB 88, 234(2015)
[7]J.-C.Delvenne,R.Lambiotte,L.E.C.Rocha,NatComm 6,7366(2015)
[8]R. Lambiotte,L.Tabourier,J.-C.Delvenne,EPJB 86,320(2013)

3.4 Modularity and the spread of perturbations in complex dynamical systems

Artemy Kolchinsky
Santa Fe Institute
Alexander Gates
Indiana University
Luis M. Rocha
Indiana University

Many complex systems are modular, in that their components are organized in tightly-integrated subsystems that are weakly-coupled to one another. Modularity has been argued to be important for evolvability, state-space exploration, subsystem specialization, and many other important functions. The problem of decomposing a system into weakly-coupled modules, which has been studied extensively in graphs, is here considered in the domain of multivariate dynamics, a commonly-used framework for modeling complex physical, biological and social systems. We propose to decompose dynamical systems using the idea that modules constrain the spread of localized perturbations. We find partitions of system variables that maximize a novel measure called 'perturbation modularity', defined as the auto-covariance of a coarse-grained description of perturbed trajectories. Our approach effectively separates the fast intra-modular from the slow inter-modular dynamics of perturbation spreading (in this respect, it is a generalization of the Markov stability method of community detection). Perturbation modularity can capture variation of modular organization across different system states, time scales, and in response to different kinds of perturbations. We argue that our approach offers a principled alternative to detecting graph communities in networks of statistical dependency between system variables (e.g. 'relevance networks', 'functional networks', and other networks based on correlation or information-transfer measures). Using coupled logistic maps, we demonstrate that the method uncovers hierarchical modular organization encoded in a system's coupling matrix. Additionally, we use it to identify the onset of self-organized modularity in certain parameter regimes of homogeneously-coupled map lattices (originally popularized by Kaneko). Our approach offers a powerful and novel tool for exploring the modular organization of complex dynamical systems.

3.5 Impact of Network Structure on Knowledge Diffusion between Groups of OSS Developers

Stefan Kambiz Behfar
BETA — CNRS, University of Strasbourg
Ekaterina Turkina
Department of international business, HEC Montreal
Thierry Burger-Helmchen
BETA — CNRS, University of Strasbourg

There have been studies in the area of knowledge management where network analysis has been used to model the knowledge transfer via weak ties between organization's units which are themselves strongly tied inside (e.g. Hansen 1999). Some studies also investigated knowledge management among open-source-software (OSS) developers and discussed factors influencing knowledge transfer within development teams (Joshi and Sarker 2006). Joshi et al Ojha (2005) also discussed knowledge sharing between team members based on similarity-attraction paradigm. At the same time, we step forward to distinguish knowledge transfer within groups as opposed to knowledge transfer between groups; where relevant projects are bundled into separate strongly intra-connected groups; whereas irrelevant groups are weakly inter-connected. In knowledge management literature there is a trade-off between sparse network structures (Burt 2000, 2002) versus dense network structure (Walker et al 1997, Coleman 1988). It is argued that the former facilitates diffusion and generation of ideas among clusters, while the latter impacts on implementation of idea within each dense cluster. To our best knowledge, there has been no study to investigate the relationship between dense and sparse network structures. At the same time we propose that knowledge transfer within dense groups influence on knowledge transfer between sparse groups. To prove our hypothesis, we use a complex network of open source software (OSS) as the domain of interest, where developers represent nodes and two developers contributing to a task represent a network tie. Developers contributing to tasks in groups other than their own can explore novel ideas, whereas developers contributing to tasks inside groups exploit ideas to improve those projects.

3.6 Automatic identification of relevant concepts in scientific publications

Andrea Martini
École Polytechnique Fédérale de Lausanne (EPFL)
Alessio Cardillo
École Polytechnique Fédérale de Lausanne (EPFL)
Paolo De Los Rios
École Polytechnique Fédérale de Lausanne (EPFL)

In recent years, the increasing availability of publication records has attracted the attention of the scientific community. In particular, many efforts have been devoted to the study of the organization and evolution of science by exploiting the textual information extracted from the title and abstract of the articles. However, lesser attention has been devoted to the core of the article, i.e., its body. The access to the entire text, instead, paves the way to a better comprehension of the relations of similarity between articles. In the present work, concepts are extracted from the body of the scientific articles available on the ScienceWISE platform, and are used to build a network of similarity between articles. The resulting weighted network possesses a considerably high edge density, spoiling any attempt of associating communities of papers to specific topics. This happens because not all the concepts inside an article are truly informative and, even worse, they may not be useful to discriminate articles with different contents. Moreover, the presence of "generic concepts" with a loose meaning implies that a considerable amount of connections is made up by spurious similarities. To eliminate generic concepts, we introduce a method to evaluate the concepts' relevance according to an information-theoretic approach. The significance of a concept c is defined in terms of the distance between its maximum entropy distribution, $S_{\max}$, and the empirical one, S_c, calculated using the frequency of occurrence inside papers. By evaluation such distance, generic concepts are automatically identified as those with an entropy closer to the maximum. The progressive removal of generic concepts retaining only the "meaningful" ones has a twofold effect: it decreases sensibly the density of the network and reinforce meaningful relations. By applying different "filtering thresholds", we unveil a refined topical organization of science in a coarse-grained way.

3.7 The dynamics of community sentiments on Twitter

Danica Vukadinovic Greetham
Billie Joe Charlton
Colin Singleton

We study a large evolving network obtained from Twitter created by a sample of users @-mentioning each other. We find that people who have potentially the largest communication reach (according to a dynamic centrality measure) use sentiment differently than the average user: for example they use positive sentiment more often and negative sentiment less often. Furthermore, we use several algorithms for community detection based on structure of the network and users' sentiment levels to identify several communities. These communities are structurally stable over a period of months. Their sentiment levels are also stable, and sudden changes in daily community sentiment in most cases can be traced to external events affecting the community. Based on our findings, we create and calibrate a simple agent-based model that is capable of reproducing measures of emotive response comparable to those obtained from the observed data.

3.8 Probabilistic and flux-based analysis of metabolic graphs

Mariano Beguerisse Diaz
University of Oxford
Mauricio Barahona
Imperial College London
Gabriel Bosque Chacón
Universidad Politécnica de Valencia
Diego Oyarzún
Imperial College London
Jesús Picó
Universidad Politécnica de Valencia

We present a framework for the construction and analysis of directed metabolic reaction graphs that can be tailored to reflect different environmental conditions. In the absence of information about the environmental context, we propose a Probabilistic Flux Reaction Graph (PRG) in which the weight of a connection between two reactions is the probability that a randomly chosen metabolite is produced by the source and consumed by the target. Using context-dependent flux distributions from Flux Balance Analysis (FBA), we produce a Flux-Balance Graph (FBG) with weighted links representing the amount of metabolite flowing from a source reaction to a target reaction per unit time. The PRG and FBG graphs are analyzed with tools from network theory to reveal salient features of metabolite flows in each biological context. We illustrate our approach with the directed network of the central carbon metabolism of Escherichia coli, and study its properties in four relevant biological scenarios. Our results show that both flow and network structure depend drastically on the environment: graphs produced from the same metabolic model in different contexts have different edges, components, and flow communities, capturing the biological re-routing of metabolic flows inside the cell. By integrating probabilistic and FBA-based analysis with tools from network science, our results provide a framework to interrogate cellular metabolism beyond standard pathway descriptions that are blind to the environmental context.