SOE 4: Computational Social Science and Data Science II

Time: Monday 12:00–12:45 Location: H17

SOE 4.1 Mon 12:00 H17

Reconstructing networks with unknown and heterogeneous errors — •Tiago Peixoto — University of Bath, UK

The vast majority of network data sets contain errors and omissions, although this fact is rarely incorporated in traditional network analysis. Recently, an increasing effort has been made to fill this methodological gap by developing network-reconstruction approaches based on Bayesian inference. These approaches, however, rely on assumptions of uniform error rates and on direct estimations of the existence of each edge via repeated measurements, something that is currently unavailable for the majority of network data. Here, we develop a Bayesian reconstruction approach that lifts these limitations by allowing for not only heterogeneous errors, but also for single edge measurements without direct error estimates. Our approach works by coupling the inference approach with structured generative network models, which enable the correlations between edges to be used as reliable uncertainty estimates. Although our approach is general, we focus on the stochastic block model as the basic generative process, from which efficient nonparametric inference can be performed and yields a principled method to infer hierarchical community structure from noisy data. We demonstrate the efficacy of our approach with a variety of empirical and artificial networks.

SOE 4.2 Mon 12:15 H17

There is more in your Data! - Analysis of Complex, Interconnected Datasets using Semantic Data Modeling — Timm Fitschen[1,2], •Alexander Schlemmer[1,3], Henrik tom Wörden[1,2], Daniel Hornung[1], Ulrich Parlitz[1,2,3], and Stefan Luther[1,2,3,4,5] — [1]Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany — [2]Institute for Nonlinear Dynamics, Georg-August-Universität, Göttingen, Germany — [3]German Center for Cardiovascular Research (DZHK), partner site Göttingen, Germany — [4]Institute of Pharmacology and Toxicology, University Medical Center Göttingen, Göttingen, Germany — [5]Department of Physics and Department of Bioengineering, Northeastern University, Boston, USA

An omnipresent challenge in the analysis of complex, heterogeneous data sets is the storage, systematic retrieval and processing of interconnected data. This especially holds for interdisciplinary fields where non-standard and rapidly evolving analysis algorithms are employed. Using a semantic data model implemented in the open source software CaosDB (https://arxiv.org/abs/1801.07653), we demonstrate in case examples from cardiac research how data from experiments and simulations can be efficiently managed and processed from data acquisition to the final publication. During this data life cycle all relevant information including metadata, analysis results, documentation and software information will be stored and linked together with the raw data. This procedure guarantees that data, results and publications are documented, findable and reproducible.

SOE 4.3 Mon 12:30 H17

Big Data and Machine Learning in Astrophysics — •Karl Mannheim — Universität Würzburg, Lehrstuhl für Astronomie

Next-generation observatories such as SKA will produce data at a rate higher than can be analyzed by human scientists. Therefore, fast analysis methods such as machine learning will play a major role in the near future. Methods applied to offline data target the extraction of scientifically relevant data from the raw data, e.g. by classifying objects or recognizing morphological patterns in images or in the spectral domain. Methods applied to online data can feedback with the data aquisition system to optimize the detector performance or reduce the data volume by filtering out unusable low-quality data. Utilizing their full potential requires to advance machine learning algorithms from finding correlations to allowing causal inferences. Users and developers of ML methods from all branches of physics, astronomy, or computers science are encouraged to discuss and exchange ideas in the new DPG working group AKPIK.