

THE ROLE OF GENERATIVE MODELS IN DATA-DRIVEN APPROACHES TO COMPLEX SOCIAL SYSTEMS



A TALK BY

MARTIN GERLACH

Postdoctoral Researcher

Department of Chemical and Biological Engineering, Northwestern University

TUESDAY, MAR 19, 11:00 A.M. | NÁDOR U. 11., ROOM 616

ABSTRACT | The increasing degree of digitization allows for new approaches for the quantitative analysis of complex social systems. Therefore, one of the major computational challenges is to develop data-driven models to extract useful knowledge from these datasets. Presenting two examples from my current research, the aim of this talk is to show how assumptions (often implicitly contained) in the model as a data-generating process affect our conclusions and can lead to severe biases. First, I will focus on statistical laws which are believed to capture universal macroscopic regularities in human behavior. Recently, however, the empirical support of these laws has been heavily questioned. Here, I will discuss how the presence of correlations -- a typical scenario for data from complex systems -- can lead to wrong conclusions about the validity of these laws when using standard maximum likelihood recipes. Second, I will focus on topic models, which have become a standard tool for organizing large collections of unstructured texts. Despite their success and numerous applications in computational social science, topic models are known to suffer from severe conceptual and practical problems, such as discrepancies with statistical properties of real texts, and the inability to properly choose the number of topics. I will present an alternative approach in the framework of community detection in complex networks using stochastic block models which leads to a more versatile and principled framework for topic modeling. I will discuss how the connection between topic models and complex networks opens the possibility of cross-fertilization to develop new computational approaches to find large-scale structures even beyond language.

BIO | Martin Gerlach is a Postdoctoral Fellow at Northwestern University in the Department of Chemical and Biological Engineering. Prior, he received his Ph.D. in Physics from the Max Planck Institute for the Physics of Complex Systems. His research aims to develop and apply mathematical tools from statistical physics, complex networks, and machine learning to gain deeper insights into the dynamics of complex social systems. His main focus is on scaling laws in complex systems, community detection and topic modeling, and the evolution of scientific disciplines.