WiSe 2021 Syllabus of the MSc Course

Statistical Network Analysis

Univ.-Prof. Dr. Ingo Scholtes
Lehrstuhl Informatik XV - Machine Learning for Complex Networks
Julius-Maximilians-Universität Würzburg

2 + 2 SWS (5 ECTS)

Executive Summary

Networks matter! This holds for technical infrastructures like communication or transportation networks, for information systems and social media in the World Wide Web, but also for various social, economic and biological systems. What can we learn from data that capture the interaction topology of such complex systems? What is the role of individual nodes and how can we discover significant patterns in the structure of networks? How do these structures influence dynamical process like diffusion or the spreading of epidemics? Which are the most influential actors in a social network? And how can we analyse time series data on systems with dynamic network topologies?

This course will equip participants with statistical analysis techniques that are needed to answer such questions based on network data across different disciplines. The course will show how networked systems can be modelled and how patterns in their topology can be characterised quantitatively. Students will understand how the topology of networks shapes dynamical processes, how statistical characteristics influence the robustness of systems, and how complex macroscopic features emerge from simple processes. The course combines a series of lectures – which introduce theoretical concepts in statistical network analysis – with weekly exercises that show how we can apply them in practical network analysis tasks. The course material consists of annotated slides for lectures, lecture videos, and `jupyter` notebooks. Students can apply and deepen their knowledge through weekly exercise sheets. The successful completion of the course requires to pass a final written exam.

Chapter I: Introduction to Network Science and Graph Theory

The first chapter of our course motivates the growing need for network analysis techniques in science, industry, and society. We introduce graph-theoretic foundations of network analysis and explain how we can mathematically analyse patterns in network topologies.

L01 What is Network Science?

We give an overview of the course and introduce interdisciplinary applications of network analysis and modeling.

- Course overview and administrative issues
- The role of networks in complex systems
- Exemplary applications of network science
- Statistical characterization of large networks

L02 Graph-theoretic foundations

We learn how to mathematically represent networks as graphs, how to compute the importance of nodes, and how to quantify cluster patterns.

- Graph theory primer
- Centrality measures
- Clustering coefficient
- Modularity and communities

Chapter II: Statistical Ensembles of Networks

The second chapter introduces the ensemble perspective on complex networks and explains analytical techniques that enable us to make strong statements about macroscopic system qualities like connectedness, diameter, or robustness based on simple aggregate statistics.

L03 Network ensembles

We introduce basic statistical ensembles of random graphs and derive expressions for the probability and degree distribution of microstates.

- Statistical ensembles: physics vs. networks
- Erdös-Renyi random graph model
- Microstate probabilities
- Degree distribution of random graphs

L04 Small-world networks

We compare the diameter and clustering coefficient of random graphs with real social networks. We introduce the small-world property and explain how it affects the navigability of networks.

- Diameter and clustering in random graphs
- Small-world effect in empirical networks
- Navigability and funneling
- Watts-Strogatz model

L05 Degree-based ensembles

We introduce the configuration model and show how we can make statements about the macroscopic properties of a network if we only know its degree distribution.

- Micro- and macro-canonical ensembles
- Molloy-Reed configuration model
- Generating functions of network ensembles
- Properties of generating functions

L06 Generating function analysis

We use generating functions to explain the friendship paradox in social networks. We further use generating functions to derive the critical point at which a giant connected component emerges in random networks.

- Generating functions in network analysis
- The friendship paradox
- Emergence of a giant connected component
- Molloy-Reed criterion in random networks

L07 Network robustness

We use generating functions to analyse the robustness of networks against random failures. We introduce scale-free networks and highlight fallacies of applying theoretical findings to real systems.

- Robustness: Random failures vs. attacks
- Scale-free networks
- Limitations of ensemble-based approaches
- Example: AS-level Internet topology

L08 Growing Networks

We introduce a simple stochastic growth model for networks and analytically show under which conditions it gives rise to scale-free networks.

- Network Growth Model
- Master Equation Analysis
- Preferential Attachment
- Copying Model

Chapter III: Dynamical Processes in Networks

The third chapter addresses the stochastic modelling of linear dynamical processes in networks. We show how we can use Markov chains to model random walks and diffusion and how spectral properties allow us to predict the evolution of processes.

L09 Random walks in networks

We explain how we can use Markov chains to model diffusion in networks and show how random walks can be used to identify important nodes.

- Modelling dynamical processes in networks
- Random walks as model for diffusion
- Markov chain convergence theorem
- Feedback centrality measures

L10 Spectral analysis

We show how we can predict evolution of dynamical processes based on eigenvalues and eigenvectors of matrix representations of networks.

- Continuous-time diffusion and Laplacians
- Eigenvalue spectrum of transition matrices
- Graph Laplacians and algebraic connectivity
- Fiedler vector and spectral partitioning

Chapter IV: Learning from Network Data

The final chapter of our course introduces statistical learning techniques for data on networked systems. We show how the ensemble perspective helps us to detect clusters and how temporal characteristics influence the results of network analysis.

L11 Stochastic block model

We learn how network ensembles can be used for statistical inference and learning tasks. We illustrate the use of ensembles to detect clusters in networks.

- Machine learning in networks
- Likelihood-based inference in network models
- Stochastic block model
- Likelihood maximisation and overfitting

L12 Flow compression

We introduce information-theoretic methods to infer the optimal number of clusters in a network based on the stochastic block model and flow compression.

- Likelihood maximisation and overfitting
- Minimum description length principle
- Huffman code and source coding theorem
- Flow Compression with the MapEquation

L13 Time Series Data on Networks

We demonstrate that the arrow of time can invalidate network analysis results and show how to address this issue with temporal network analysis.

- Time Series Data on Networks
- Temporal vs. Growing Networks
- Causal paths in Temporal Networks
- Diffusion and Spreading in Temporal Networks

L14 Beyond Simple Network Models

We show how we can use higher-order network models to improve the analysis of complex systems.

- Networks as Topological Spaces
- Higher-order network models
- Higher-order Laplacians
- Model selection and cross-validation