Autumn 2018 Syllabus of the MSc Course on

Statistical Network Analysis

Prof. Dr. Ingo Scholtes
Department of Informatics
University of Zurich

Summary

Networks matter! This holds for technical infrastructures like the Internet, 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 the topology of such complex networked systems? What is the role of individual nodes and how can we discover significant patterns in the global structure of networks? How do these structures influence dynamical process? Which are the most influential actors in a social network? And how can we analyse time series data on networks with dynamic topologies?

In this course, students get a broad overview about statistical modeling and analysis techniques that can be used to study complex networks across disciplines. The course will show how networks can be represented mathematically and how patterns in their topology can be characterized quantitatively. Students will understand how networks shape dynamical processes and how complex link topologies emerge from simple network formation processes. We further explore how data mining and machine learning techniques can be applied to extract knowledge from complex relational data on technical, social, and economic systems.

The accompanying exercises consist of computer simulations and real-world data analysis tasks from different disciplinary contexts that should be solved using `python`. An introductory tutorial, sample programs, and code skeletons will be provided during the course, so no prior knowledge of `python` is required. During the weekly exercise sessions, students are expected to present their solutions, which will then be discussed. Sample solutions are provided after the exercises.

Chapter I: Introduction to Network Science

Lecture 01 – Motivation

Educational Objective: In this lecture, participants will get an overview of the course and will see interdisciplinary applications of network analysis and modeling.

- Course overview and administrative issues
- Motivation: the role of network topology in complex systems
- Examples for interdisciplinary applications of network science
- Statistical characterization of large complex networks from technology, society, life sciences, and economics

Exercise 01: Introduction to `python`, `jupyter`, and `pathpy`

Lecture 02 – Network analysis primer

Educational Objective: In this lecture, students will learn how to mathematically represent complex networks and how to quantitatively analyse the importance of nodes.

- Basic definitions: graph, network, adjacency matrix, path, cut, degree
- Importance of nodes: betweenness, closeness and degree centrality
- Modules and clusters: clustering coefficient and modularity
- Hands-on analysis: centrality analysis in a collaboration network from software engineering

Exercise 02: Analysis of empirical networks with `pathpy`

Chapter II: Statistical Ensembles of Complex Networks

Lecture 03 – Network ensembles

Educational Objective: In this lecture, participants will learn how networks can be represented and analysed from a statistical point of view.

- Graph theory vs. network science: the ensemble perspective
- The Erdös-Renyi random graph model
- Degree distribution and average degree in random graphs
- Hands-on analysis: degree distribution of coauthorship and citation networks in science

Exercise 03: Implementing and analyzing statistical ensembles in `python`

Lecture 04 – Clustering and small-world networks

Educational Objective: In this lecture, students will learn how to generate networks that reproduce the small diameter and large clustering coefficient observed in real-world social networks.

- Degree distribution, diameter and clustering coefficient of random networks
- Navigability and funneling in small-world networks
- Watts-Strogatz model: average shortest path length and clustering coefficient
- Hands-on analysis: strong and weak ties in a social contact network at a university campus

Exercise 04: The small-world phase transition in the Watts-Strogatz model

Lecture 05 – Ensembles with fixed degree distribution

Educational Objective: In this lecture, students will learn how to can make statements about the properties of a network if one only knows the distribution of node degrees.

- Ensembles with fixed distribution of degrees: Molloy-Reed model
- The generating functions framework
- Properties of generating functions
- Hands-on analysis: the friendship paradox in an online social network

Exercise 05: Analysing component size of random networks

Lecture 06 – Generating function analysis of network ensembles

Educational Objective: In this lecture, participants will learn how the framework of generating functions can be applied to study the emergence of a giant connected component in large complex networks.

- Friendship paradox and sampling biases
- Emergence of a giant connected component
- The Molloy-Reed criterion
- Hands-on analysis: component analysis in scientific coauthorship networks

Exercise 06: Random node failures in complex networks

Lecture 07 – Robustness and scale-free networks

Educational Objective: In this lecture, participants will learn what fallacies one encounters when applying findings from ensemble studies to real-world networks.

- Generating function analysis of network robustness
- Scale-free networks
- Limitations of ensemble-based approaches
- Hands-on analysis: robustness of the AS-level Internet topology

Exercise 07: Finite-size effects in scale-free networks

Chapter III: Dynamical Processes in Networks

Lecture 08 – Random walks and diffusion in networks

Educational Objective: In this lecture, students will learn how we can model diffusion processes in complex networks by means of random walks.

- Modelling dynamical processes in networks
- Random walks as model for diffusion processes
- Markov chain convergence theorem
- Hands-on analysis: modelling diffusion in the London Tube system

Exercise 08: Simulating diffusion processes in `python` and `pathpy`

Lecture 09 – Spectral analysis of complex networks

Educational Objective: In this lecture, students will learn how spectral properties of matrix representations of networks can be used to analyse network efficiency, find central nodes and detect cluster structures.

- Feedback centrality: eigenvector centrality and PageRank
- Diffusion speed: Eigenvalue gap of transition matrices
- Graph Laplacians: Alegbraic connectivity and spectral partitioning
- Hands-on analysis: PageRank in the Wikipedia hyperlink graph

Exercise 09: Spectral analysis using `python` and `scipy`

Chapter IV: Learning in Network Data

Lecture 10 – Statistical inference: stochastic block model

Educational Objective: In this lecture, students will learn how the ensemble perspective on complex networks can be used to discover community patterns in network data.

- Introduction to Machine Learning
- Likelihood-based inference in network models
- Stochastic block model and community detection
- Hands-on analysis: community structures and overfitting in social networks

Exercise 10: Community detection using the stochastic block model

Lecture 11 – Model selection: minimizing description length

Educational Objective: In this lecture, students will learn how information-theoretic concepts can be applied to avoid an overfitting in the statistical analysis of patterns in network data.

- Shannon Entropy of network ensembles
- Stochastic block model: minimising description length
- Flow compression: the InfoMap algorithm
- Hands-on analysis: community detection in the Wikipedia hyperlink graph

Exercise 11: Detecting optimal community structures with flow compression in `pathpy`

Lecture 12 – Exponential Random Graph Models

Educational Objective: In this lecture, students will learn how Exponential Random Graph Models can be used to test hypothesis in social networks.

- Exponential Random Graph Models
- Statistical Inference with ERGMs
- Monte-Carlo Markov Chain Maximum Likelihood Estimation
- Hands-on analysis: ERGM analysis of the Florentine families network

Exercise 12: Inference of ERGMs in `python`

Chapter V: Temporal Network Data

Lecture 13 – Modeling growing networks

Educational Objective: In this lecture, students will learn how to model feedback phenomena in the stochastic growth of complex network structures over time.

- Stochastic models of growing networks
- Growth of random networks: uniform attachment
- Feedback in network growth: preferential attachment
- Hands-on analysis: modelling feedback phenomena in citation networks

Exercise 13: Simulating network growth models in `python`

Lecture 14 – Analysis of time series network data

Educational Objective: In this lecture, students will learn that the arrow of time introduces an additional dimension of complexity that needs to be incorporated into the analysis of temporal data on networks with dynamic topologies.

- Temporal networks: causal paths and temporal centralities
- Markovian vs. non-Markovian models of causal paths temporal networks
- Representation learning: Learning optimal models for causal paths
- Hands-on analysis: causal paths in a dynamic communication network in software engineering

Exercise 14: Temporal network analysis and representation learning in `pathpy`