

DNDS 6016: Statistical Methods in Network Science 2

Course syllabus

Instructor:	Prof. Gerardo Iñiguez
Term:	Winter 2019/20
Course level:	PhD, MA
Credits:	2
Host unit:	Department of Network and Data Science (DNDS)
E-learning:	To be announced
Office hours:	Upon agreement

Course Description

The increasing volume and nature of big datasets in business, economics, social and political sciences call for more complex and sophisticated mathematical and data mining tools. The complex systems monitored by big databases are successfully described in terms of networks. In this course we will present and discuss mathematical and data mining tools used to characterize large empirical or model networks. Large datasets will be computationally investigated and the limits of the used algorithms will be discussed. The assessment of the statistical validity of the observed results will be analyzed and, when possible, quantitatively evaluated. Besides the mathematical theory, the course will have a practical approach with homework and the development of a project. The project will involve data analysis and computer simulations in Python (with e.g. Jupyter notebooks).

Learning outcomes

By successfully completing the course the students will be able to:

- Learn how to perform data analysis and use statistical methods in the investigation of networks.
- Evaluate the statistical reliability of empirical estimations against an appropriate null hypothesis.
- Learn how to make use of large sets of data for investigating networks observed in the social sciences and other fields.
- Perform empirical analyses and statistical validation of large datasets obtained from the Internet or from other business and scientific sources.

Assessment

(1) Assessment type 1 (50% of the final grade). Attendance in at least 80% of classes, active cooperation, and homework: Students will get home assignments potentially consisting of statistical analyses, simple problems or data processing, which they will have to complete individually and submit electronically.

(2) Assessment type 2 (50% of the final grade). In the final project work students will have to individually perform a research project involving the investigation of a large network. The investigation and characterization of the network might include one or more of the following aspects: (i) developing a network model, (ii) running network simulations, (iii) comparing network metrics with null hypotheses, (iv) critically analyzing results of a large scale empirical investigation. Students will have to prepare a project presentation and a written report.

Prerequisites

- DNDS 6011: Statistical Methods in Network Science 1.
- Proven proficiency with Python.
- Knowledge of fundamental network science concepts.
- Basic skills in probability, statistics, linear algebra, and calculus.

The instructor holds no responsibility in case students do not satisfy the prerequisites and need to drop the course.

Tentative course topics

1 Introduction

Overview of the course. Basic concepts that will be treated in the course. Summary of prerequisite course (Statistical Methods in Network Science 1): Network science, probability and statistics.

2 Sampling and estimation in networks

Basic elements of statistical sampling theory. Induced and incident subgraph sampling. Star and snowball sampling. Estimation of original network graphs. Examples from social networks, Internet, and species detection.

3 Statistical paradoxes on networks

Friendship and generalised friendship paradox (in coauthorship, social networks, and happiness studies). Majority illusion. Simpson's paradox.

4 Motif detection

Motifs in a directed graph. Basic motifs and triad census. Efficient algorithms to find motifs. Optimality of the triangle-finding algorithm. Shuffling methods in directed networks. Application to financial crises and human mobility.

5 Signed networks and structural balance

Signed networks. Structural balance and triadic closure. Differences between positive and negative tie networks. Shuffling methods and null models. Applications to social, alliance, and biological networks.

6 Recommendation systems and link prediction

Problem description and evaluation metrics. Local similarity indices. Global similarity indices. Applications of link prediction in online social networks. Classification of partially labeled nodes.

7 Applications of maximum matching

Examples of maximum matching in shareability networks, network control theory, and organ exchange.

8 Spreading dynamics on networks.

Epidemic spreading. Diffusion of ideas and innovations. Opinion formation, consensus and polarization. Role of network structure in spreading dynamics. Data-driven examples.

9 Guest lecture

Selected topic in statistical methods in network science by a guest lecturer. To be announced.

10 Project presentations by students

Suggested reading

- Newman, Mark. Networks: Second edition. Oxford University Press, 2018.
 - Latora, Nicosia, Russo. Complex Networks. Cambridge University Press, 2017.
 - Kolaczyk. Statistical Analysis of Network Data. Springer, 2010.
 - Easley, Kleinberg. Networks, crowds, and markets, Cambridge University Press, 2010.
 - Leskovec, Rajaraman, Ullman. Mining of Massive Datasets. Cambridge University Press
- A list of papers and online resources will be provided during class.

Further information

Further information, such as the course website, assessment deadlines, office hours, contact details, etc. will be given during the course. The instructor reserves the right to modify this syllabus as deemed necessary any time during the term. Any modifications to the syllabus will be discussed with students during a class period. Students are responsible for information given in class.