

Statistical Network Analysis

Carsten Jentsch, Jonathan Flossdorf

Business and Social Statistics
Department of Statistics
TU Dortmund University

Winter Term 2022/2023

What is a network?

A network (also called graph) is a purely formal representation consisting of a collection of **nodes** and associated **links**.

Example of nodes: persons, places, firms, countries, web pages, email addresses, electrical power stations, molecules,...

Example of links: trades, marriages, citations, letters sent, contiguity, group memberships, airline connections, friendships,...

Visualization

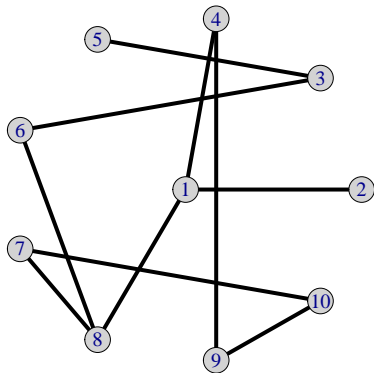


Figure: Example of a simple statistical network

Examples of Networks

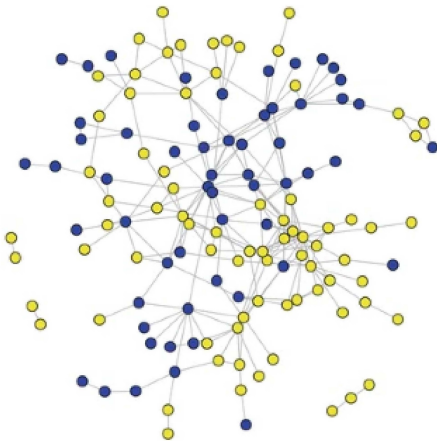


Figure: Network of interactions among proteins known to be responsible for cell communication in yeast. Yellow vertices denote proteins that are known to be involved in intracellular signaling cascades, a specific form of communication in the cell.

Examples of Networks

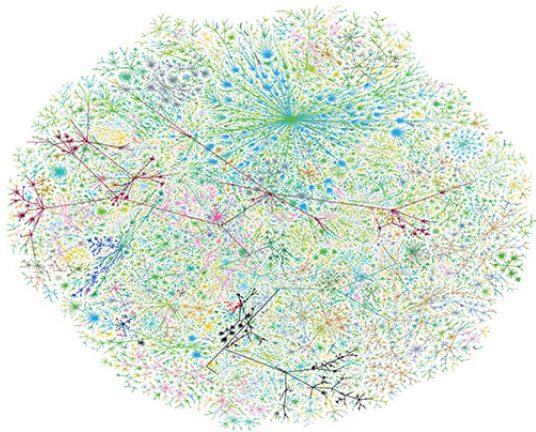


Figure: The structure of the Internet at the level of autonomous systems. The nodes are autonomous systems and the edges show the routes taken by data traveling between them.

Examples of networks

- Technological applications
 - ▶ Transportation networks
 - ▶ Energy networks
- Social networks
 - ▶ friendships
 - ▶ memberships of people in clubs or companies
 - ▶ contacts between people
- Biological networks
 - ▶ regularity behaviour among genes
 - ▶ bindings among proteins
 - ▶ epidemiological networks

Typical Questions

- What are the structural characteristics of the network?
- How much traffic is flowing across the network?
- Which nodes are the most influential ones in the network?
- Is the graph clustered (i.e. are there groups/cliques in the network)?
- Is the network behaviour as expected?
- How can we measure the network dynamics?
- How can the relationships be modeled?

Some topics that will be discussed

- Concepts for representing networks
- Components and families of (sub)graphs
- Measuring network characteristics
- Incidence matrix and Graph Laplacian
- Cohesiveness and clustering over networks
- Centrality measures
- Network models

Literature

- Kolaczyk, E. & Csardi, G. (2014): *Statistical Analysis of Network Data with R*, Springer.
- Newman, N. (2018): *Networks*, Oxford University Press.
- Salter-Townshend, M., White, A., Gollini, I., & Murphy, T. B. (2012): *Review of statistical network analysis: models, algorithms, and software*, *Statistical Analysis and Data Mining*, 5(4), 243-264.

Admin

Lectures:

- on campus (if possible), starting October 11, 2022
- mixture of slides and writing on white/black board
- one lectures per week
- Tuesday 14-16 (CDI 120)
- Moodle Room

Tutorials:

- M.Sc. Jonathan Flossdorf
- one tutorial every two weeks
- on campus (if possible)
- by-weekly exercise sheet; solutions to be handed in
- corrections if number of participants allows, otherwise 1-2 graded tests during the semester
- 40% of the total points required to be eligible for the exam

Exam:

- oral exams (if number of participants allows)

General Information:

- (2+1) course
- 4,5 ECTS
- Modules: MS 6/ MS 7 (Statistics), MD E1 (Data Science), ME 7 (Econometrics)
- course can also be attended by bachelor students