Emerging Models and Paradigms in Network Science
Part #x: Working with Network Data

Ph.D. Course at University of Calabria
27–30 May 2024

Francesco Cauteruccio

DIEM, University of Salerno

✉ fcauteruccio@unisa.it

🌐 francescocauteruccio.info

Working with graph data

- Where to find graph data?
- How to handle graph data?
- How can I visualize my graphs?
- Is there anything else out there?

Where to find graph data?

- SNAP (Stanford Network Analysis Project) datasets
  - http://snap.stanford.edu/data/index.html
  - Several kinds of networks (social, communication, citation, web graphs, road networks, etc...)
  - Each dataset is associated to a peer-reviewed paper.

- Network Repository
  - https://networkrepository.com
  - An interactive repository with several categories and hundreds of dataset for each of them.

- ARB Data
  - https://www.cs.cornell.edu/~arb/data/
  - Datasets on temporal networks, hypergraphs, trajectories, etc.

How to handle graph data?

- The mentioned repositories contain graph data in several formats
  - CSV, list of edges, binary matrices, etc...

- To handle such data, we can use existing libraries (mostly in Python or R)

- NetworkX (https://networkx.org/)
  - Python package for the creation, manipulation, and study of the structure, dynamics, and functions of networks.
  - Very intuitive, it contains a number of algorithms ready-to-use.
  - Not very useful for large-scale networks.

- HyperNetX (https://pnnl.github.io/HyperNetX/)
  - Python library for hypernetworks.

How can I visualize my graphs?

- If we are working with sufficiently small networks, being able to visualize them can help in getting a "big picture" of the system.
- Gephi (https://gephi.org/)
  - It is a visualization and exploration software for networks.
  - It also allows to perform exploratory data analysis (EDA), as well as link analysis, common social network analyses, centralities computation, etc.

Is there anything else out there?

- graph-tool (https://graph-tool.skewed.de/)
  - Python module for manipulation and statistical analysis of graphs.

- igraph (https://igraph.org/)
  - A collection of network analysis tools programmable in R, Python, C/C++ and Mathematica.

- NetworKit (https://networkit.github.io/)
  - Open-source toolkit for large-scale network analysis. It is a Python module written in C++.

- snap (https://snap.stanford.edu/snap)
  - General purpose, high performance system for analysis and manipulation of large networks. Written in C++, it also provides a Python interface.

- Cytoscape (https://cytoscape.org/)
  - Open source software platform for visualizing complex networks.

- ASH (https://github.com/GiulioRossetti/ASH)
  - ASH is a Python software package that allows the representation and analysis of dynamic hypergraphs enriched with node attributes.

Thanks for your attention!

Francesco Cauteruccio, Ph.D.
fcauteruccio@unisa.it
https://francescocauteruccio.info