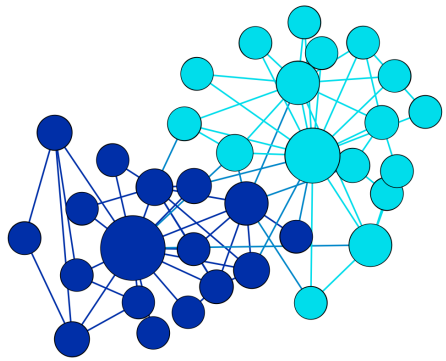
Analysis of Social and Information Networks (PRODEI040)

Pedro Ribeiro

DCC/FCUP

2021/2022



Who am I?

Name: Pedro Ribeiro
Office 1.47 (FC6 - DCC building)
Personal Website: http://www.dcc.fc.up.pt/~pribeiro/

PhD in **Computer Science**

Main research interests:

- Complex Network Analysis, Network Science, Graph Mining, Data Mining.
- Algorithms and Data Structures, Complexity.
- Parallel and Distributed Computing
- Bioinformatics Applications; Biological Networks; Social Networks

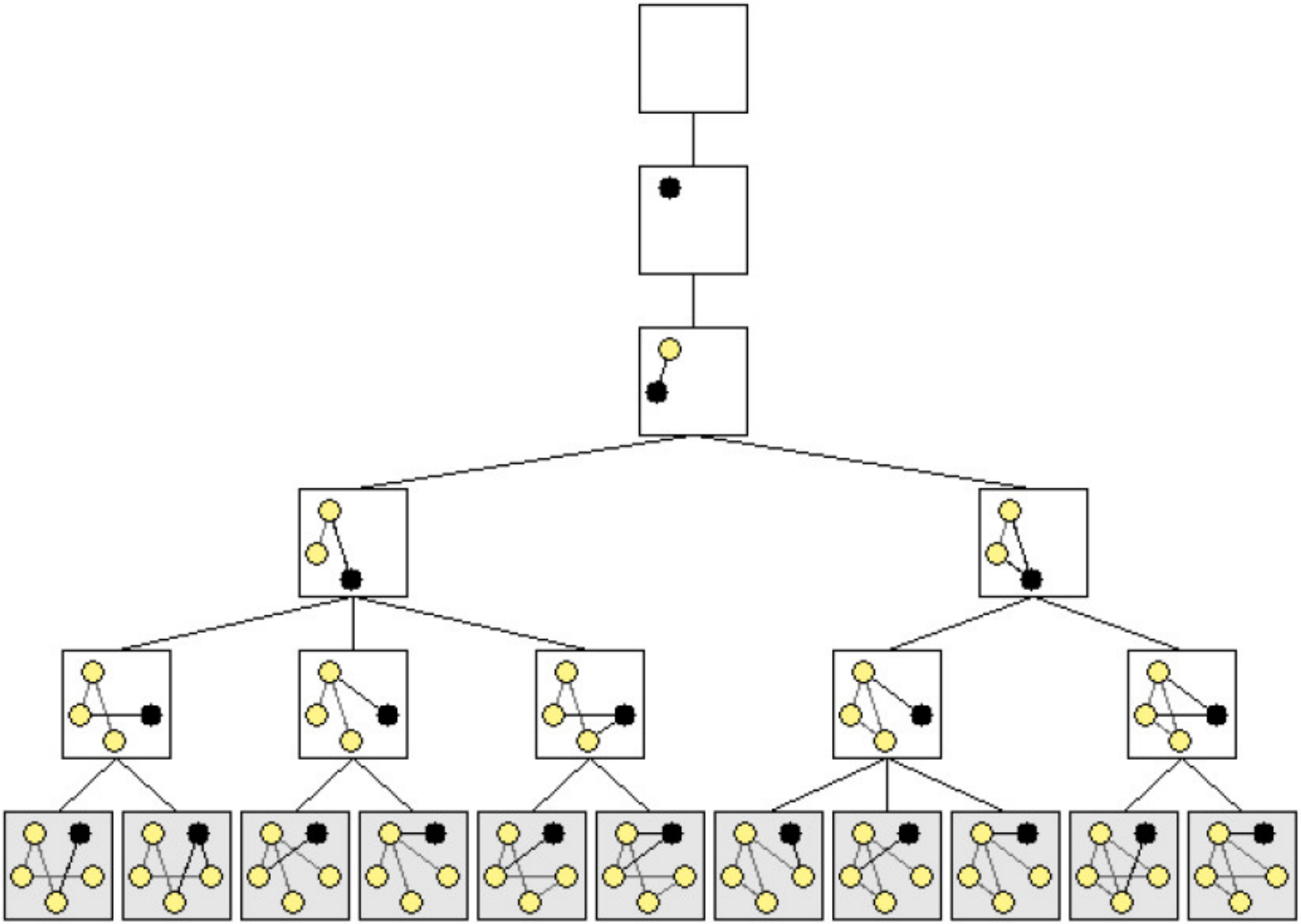
Other research interests:

- Computer Science Education and Programming Contests.
- Artificial Intelligence

PhD in Computer Science (2011)

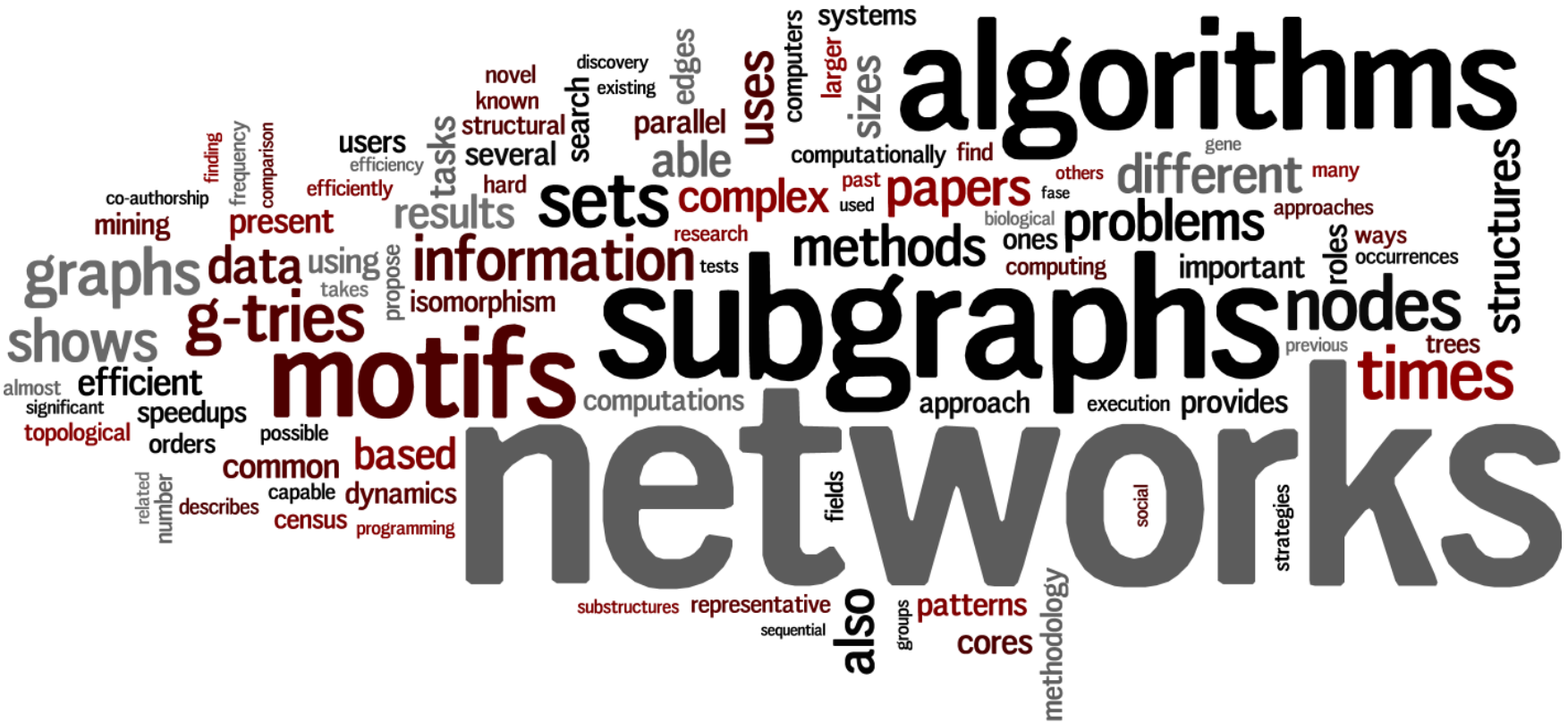
Thesis: Efficient and Scalable Algorithms for Network Motifs Discovery

PhD: G-Tries



Publications

A word cloud of my publications abstracts *(made in Dec 2015)*



http://www.dcc.fc.up.pt/~pribeiro/pubs_by_year.html

https://scholar.google.com/citations?user=DbR0sO4AAAAJ

Some of my former students

- **PhD Students**
  - Sarvenaz Choodbdar (2010-2015):
    On the Characterization and Comparison of Complex Networks
  - Miguel Araújo (2012-2017)
    Communities and Anomaly Detection in Large Edge-Labeled Graphs
  - David Aparício (2014-2020)
    Network Comparison and Node Ranking in Complex Networks
  - Jorge Silva (2016-2021)
    Towards measuring scientific impact using network science

- **MSc Students**
  - B. Pinto (20/21): *Subgraph Patterns in Colored Networks*
  - H. Branquinho (19/20): *Counting Subgraphs in Streaming Networks*
  - F. Bento (19/20): *Characterizing the Passing Networks of Football Teams*
  - L. Grácio (18/19): *From Supergraph Counting to Subgraph Generation*
  - A. Meira (18/19): *Subgraph Patterns in Multiplex Networks*
  - M. Martins (18/19): *Condensed Graphs: Towards a General Approach for Faster Subgraph Census*
  - *(...)*

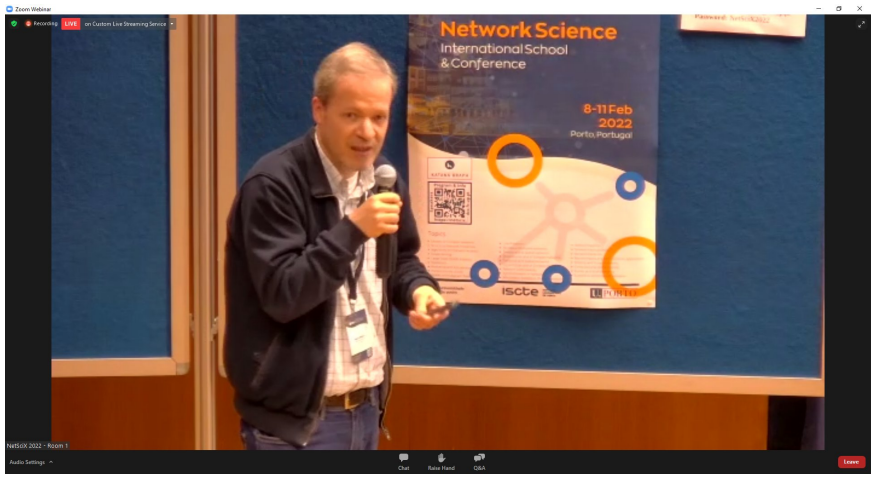
Some of my current students

- **PhD Students**
  - Vanessa Silva (since 2018):
    Analysing Time Series using Complex Networks
  - Alberto Barbosa (since 2018)
    Sports Analytics meets Network Science
  - Ahmad Naser Eddin (since 2019)
    Fraud and Anti-Money Laundering Detection using Network Science
  - Luciano Grácio (since 2020)
    Fundamental contributions on Subgraph Counting and Graph Theory
  - *(...)*

- **MSc Students**
  - Filipe Justiça: *(time series prediction using network science)*
  - José Ferreira *(spatial network motifs)*
  - Isac Novo *(graph compression and information content of subgraphs)*
  - Manuel Lamas *(characterizing music tracks using complex networks)*
  - *(...)*

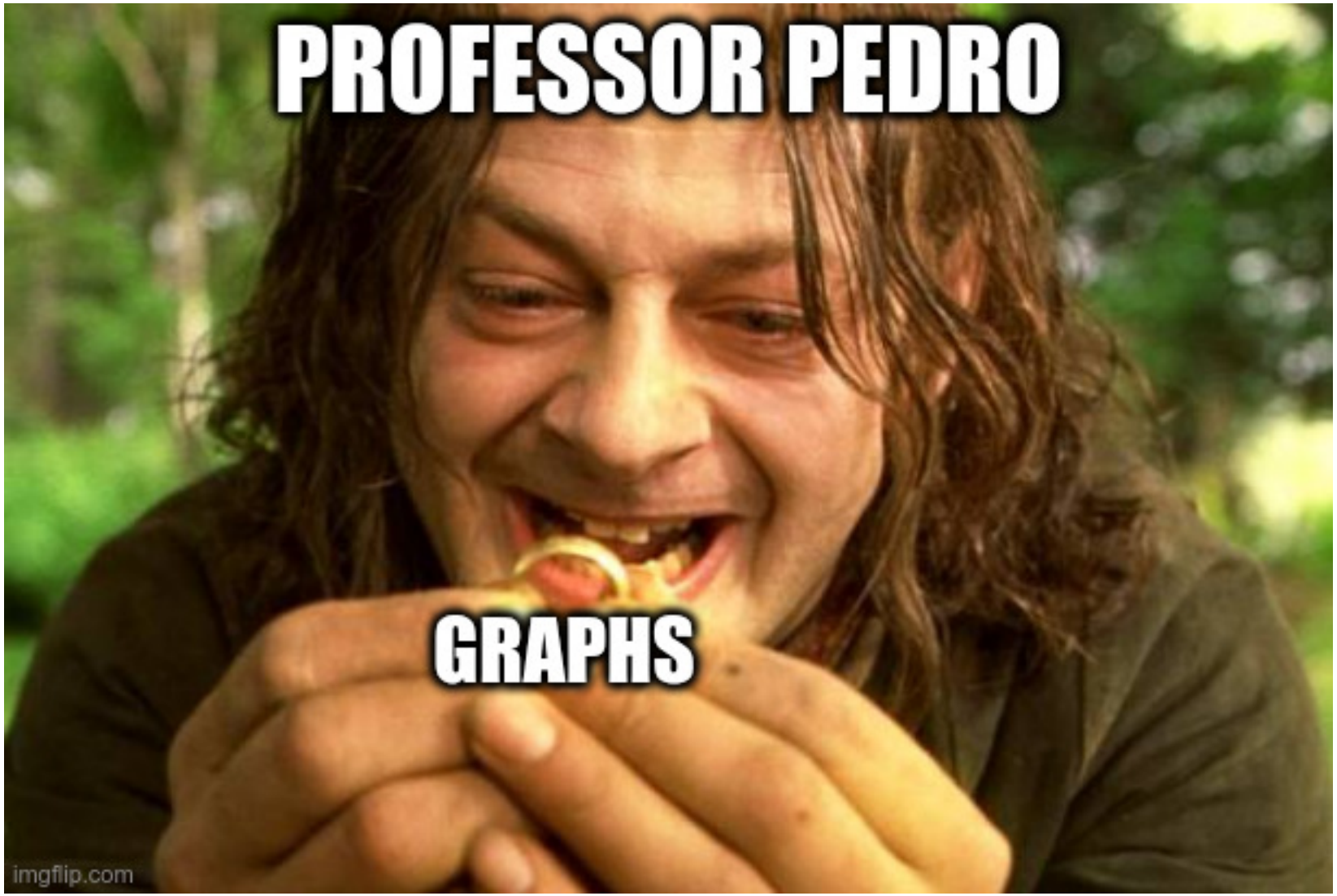
Network Science Events

http://netsci18.dcc.fc.up.pt/



https://netscix.dcc.fc.up.pt/

Graphs



General Information

- **Site:** http://www.dcc.fc.up.pt/~pribeiro/aulas/arsi2021/
- **Classes:**
  - 1h30m (online - pre-recorded videos)
  - 30m (Zoom and/or "on site" - mondays at 18:15 - non mandatory)
- **Class Organization:**
  - Theoretical Exposition: pre-recorded videos (YouTube)
  - Communication: Slack
- **Class participation:** There is no minimum enforced attendance

Evaluation

- 2 Homeworks **(40%)**
- Presentation/Reviewing an Article **(15%)**
- Project **(45%)**

There will be no final exam.

There are no minimum grades in any of the evaluation components, but failure to deliver and present the project will result on a **RFC** evaluation (missing an evaluation component).

Homeworks

2 small individual homeworks (at least 3 weeks for each) to be delivered by email.

You will apply some concepts in practice (potentially using a computer to analyze small datasets).

You can discuss with me if you have difficulties when trying to do the homework

Predicted dates:

- **1st Homework (20%)** - End of March/beginning of April
- **2nd Homework (20%)** - End of April/beginning of May

Presentation/Reviewing an Article

- This is an **individual** assignment.
- Select a recent **scientific article** (year of publication ≥ 2015) about **network science**; carefully read it and present it to the class
- You should select a **topic that interests you** and confirm with the instructor that your choice is ok. Website will have some initial pointers to help in your search for a suitable article.

- **Date of presentation:** end of semester (possibly pre-recorded?)
- **Time for each presentation:** around 15 minutes

Project

- This is a **group** assignment (maximum: 2 students)
- **Analyze a network dataset** using the concepts you learned
- You can use your own built dataset of create a new one
- Focus can be on the **analysis itself** (ex: new insights), on the **implementation** (ex: compare efficiency) or any other combination
- You are strongly encouraged to **speak with your instructor to validate your project** idea before starting your work.

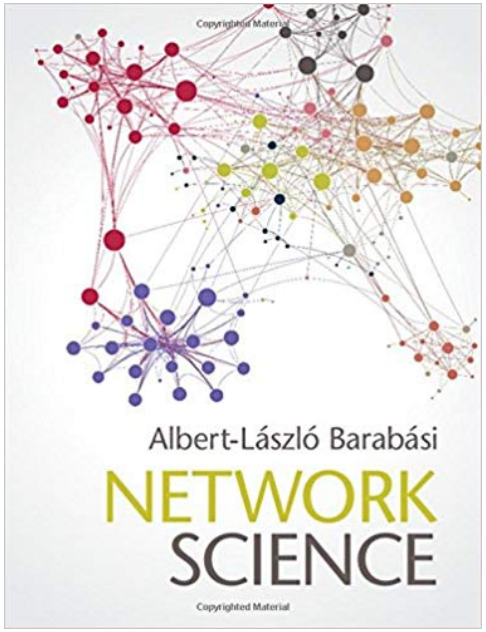
What do I need to deliver?

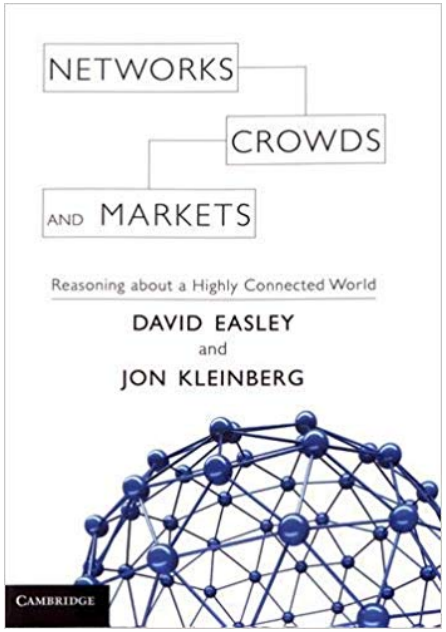
- **Written article:** 6 to 10 pages (KDD format)
  [deadline: end of classes]
- **Presentation:** 15 to 30 minutes

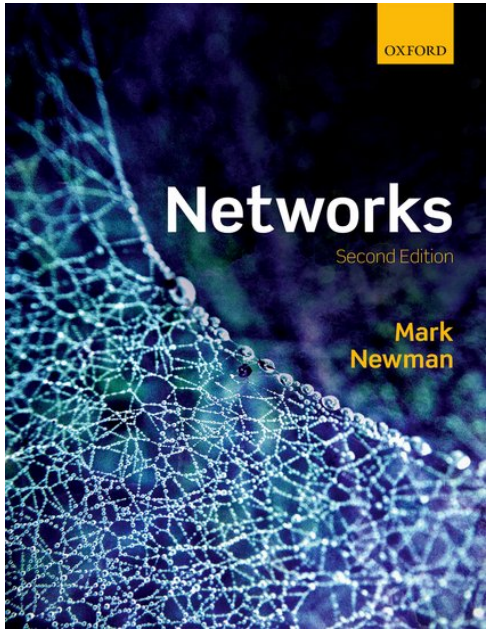
Learning outcomes in this curricular unit

- explain the **key concepts of network science** and network analysis
- apply a range of techniques for **characterizing network structure**
- define **methodologies for analyzing networks** of different fields
- demonstrate **knowledge of recent research** in the area

Recommended Books



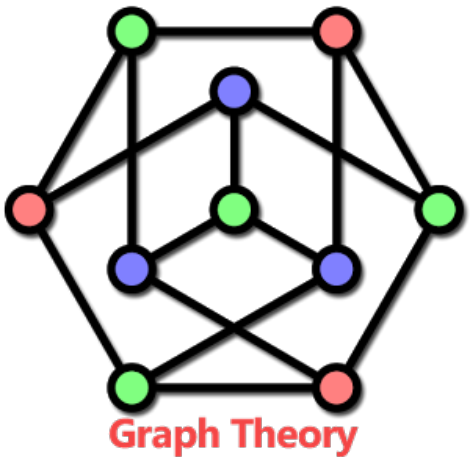




An overview of the program

Fundamentals of Network Science:

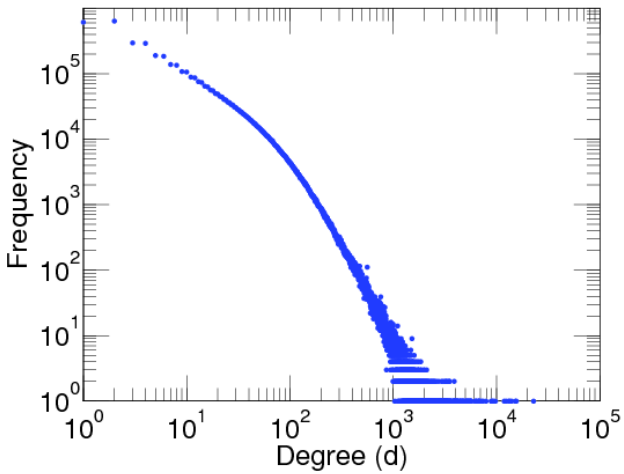
- Emergence of Network Science
- Essential Graph Theory
- Classic Graph Algorithms and Data Structures
- ...

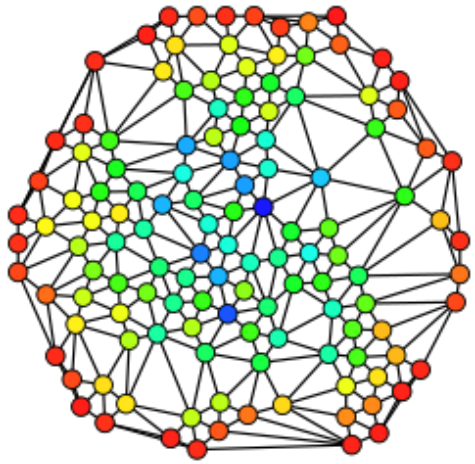


An overview of the program

Metrics and Basic Structural Properties:

- Degree distribution
- Paths and diameter
- Clustering coefficient
- Classic Node Centralities (betweenness, closeness, eigenvector, ...)
- ...

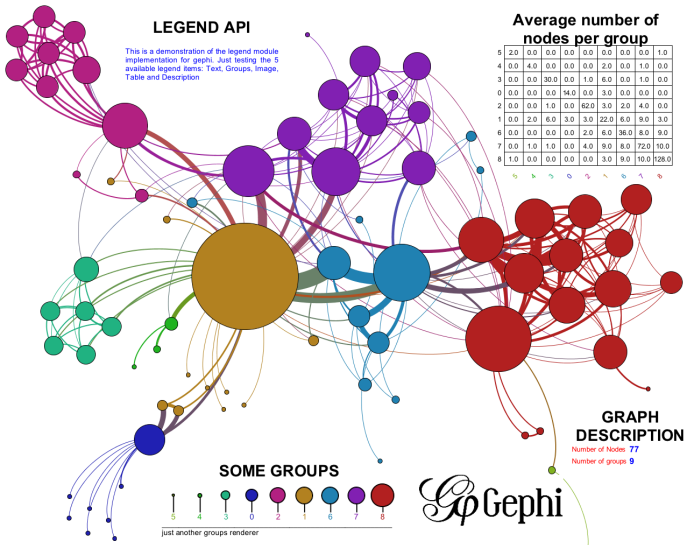


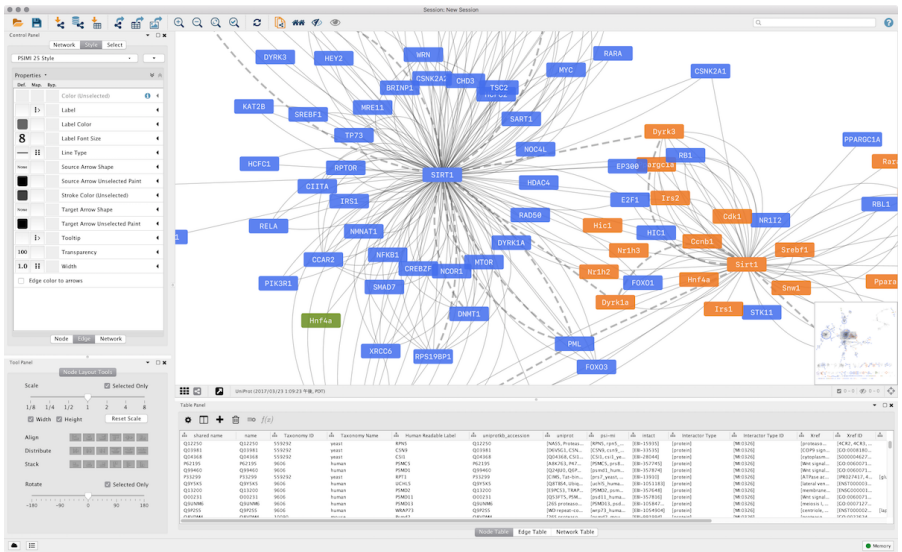


An overview of the program

Network Visualization:

- Graph Drawing
- Layout Algorithms
- Exploratory Analysis with Visualization
- ...

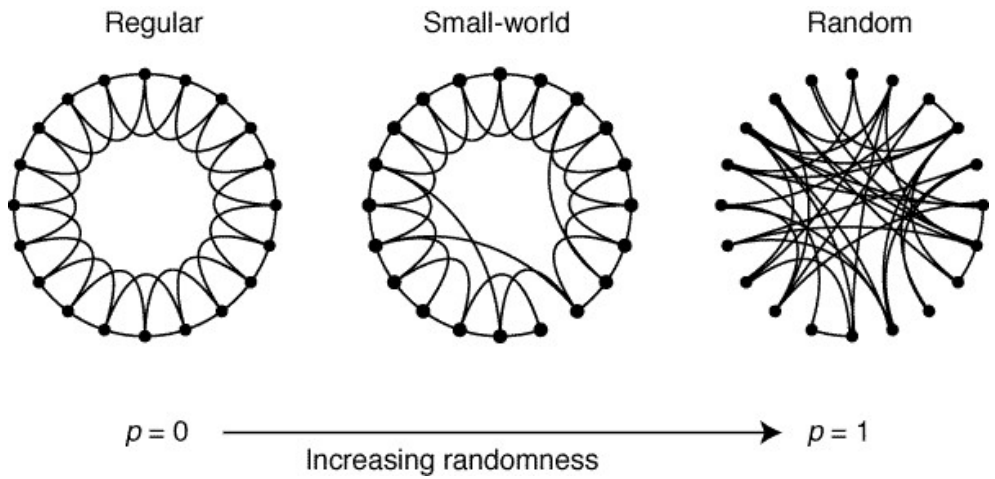




An overview of the program

Common Properties and Network Models:

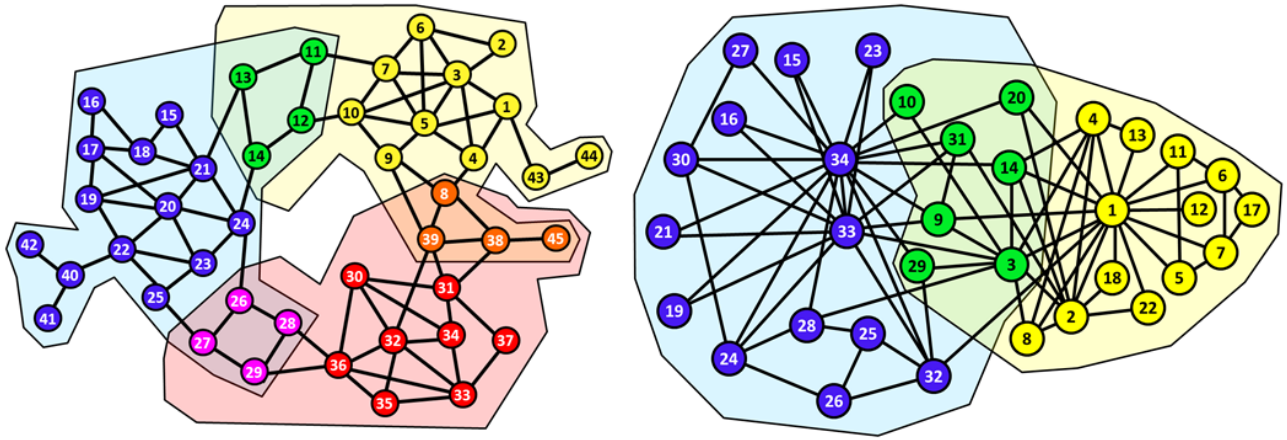
- Random networks and Erdös-Rényi model
- "small-world" property and Watts-Strogatz model
- "scale-free" property and Albert-Barabasi model
- Other models (ex: Kronecker graphs).
- ...



An overview of the program

Community Detection:

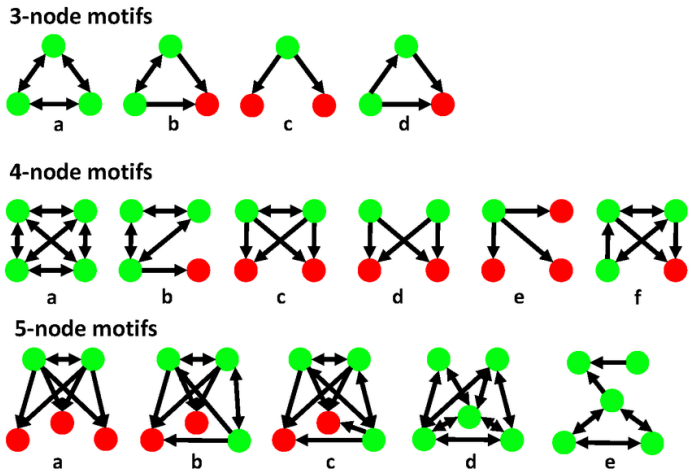
- Modularity
- Algorithms for detecting communities
- Overlapping communities and other variants
- ...

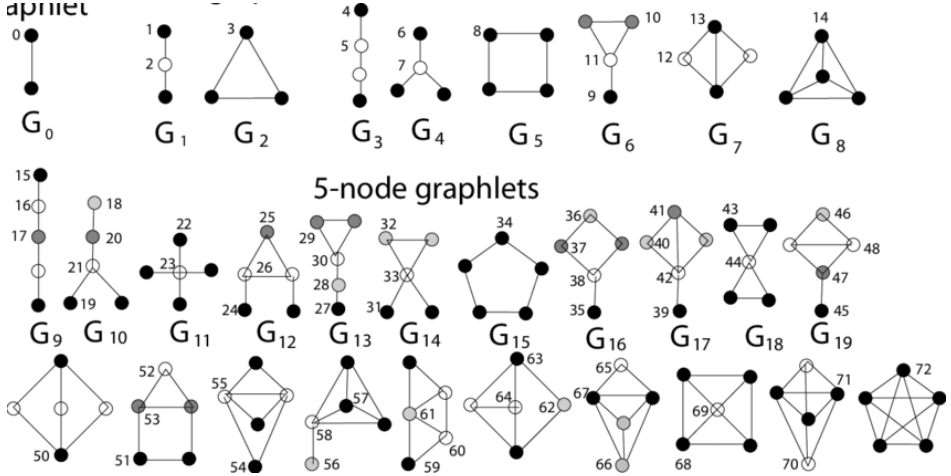


An overview of the program

Patterns and Subgraphs:

- Subgraph Census
- Network Motifs
- Graphlets
- Fingerprinting Networks
- ...

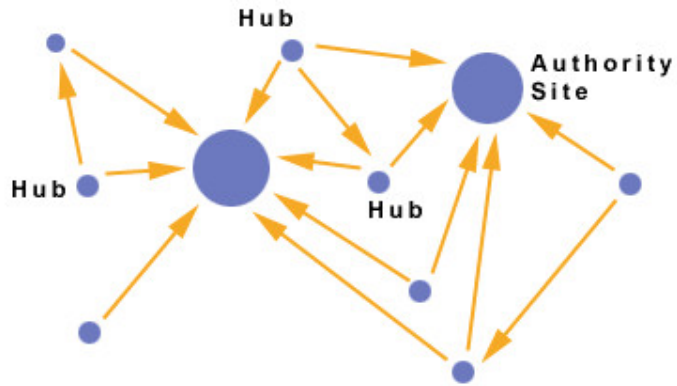




An overview of the program

Link Analysis:

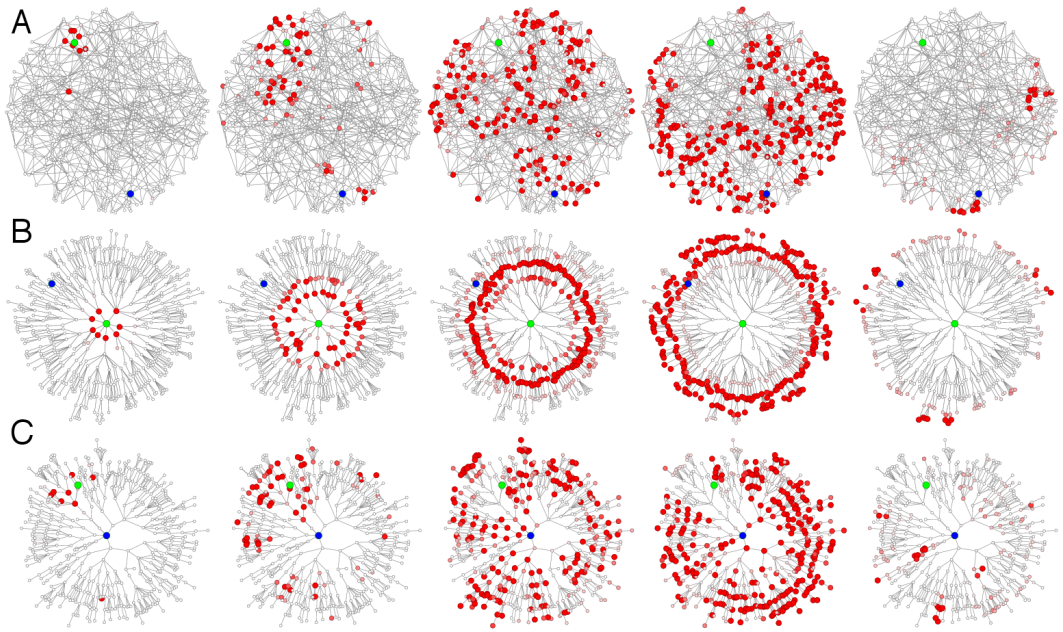
- Node Rankings
- HITS algorithm
- PageRank and other variants
- ...



An overview of the program

Propagation in networks:

- Information flow
- Epidemics
- Propagation models
- ...



An overview of the program

Other Selected Topics:

- Link Prediction
- Network Alignment
- Graph Databases
- Temporal Networks
- Multilayer networks
- Graph Neural Networks
- Graph APIs (networkx, igraph, snap, ...)
- ...