Learning from Networks

Computer Engineering

Fabio Vandin

October 2nd, 2024

Learning from Networks

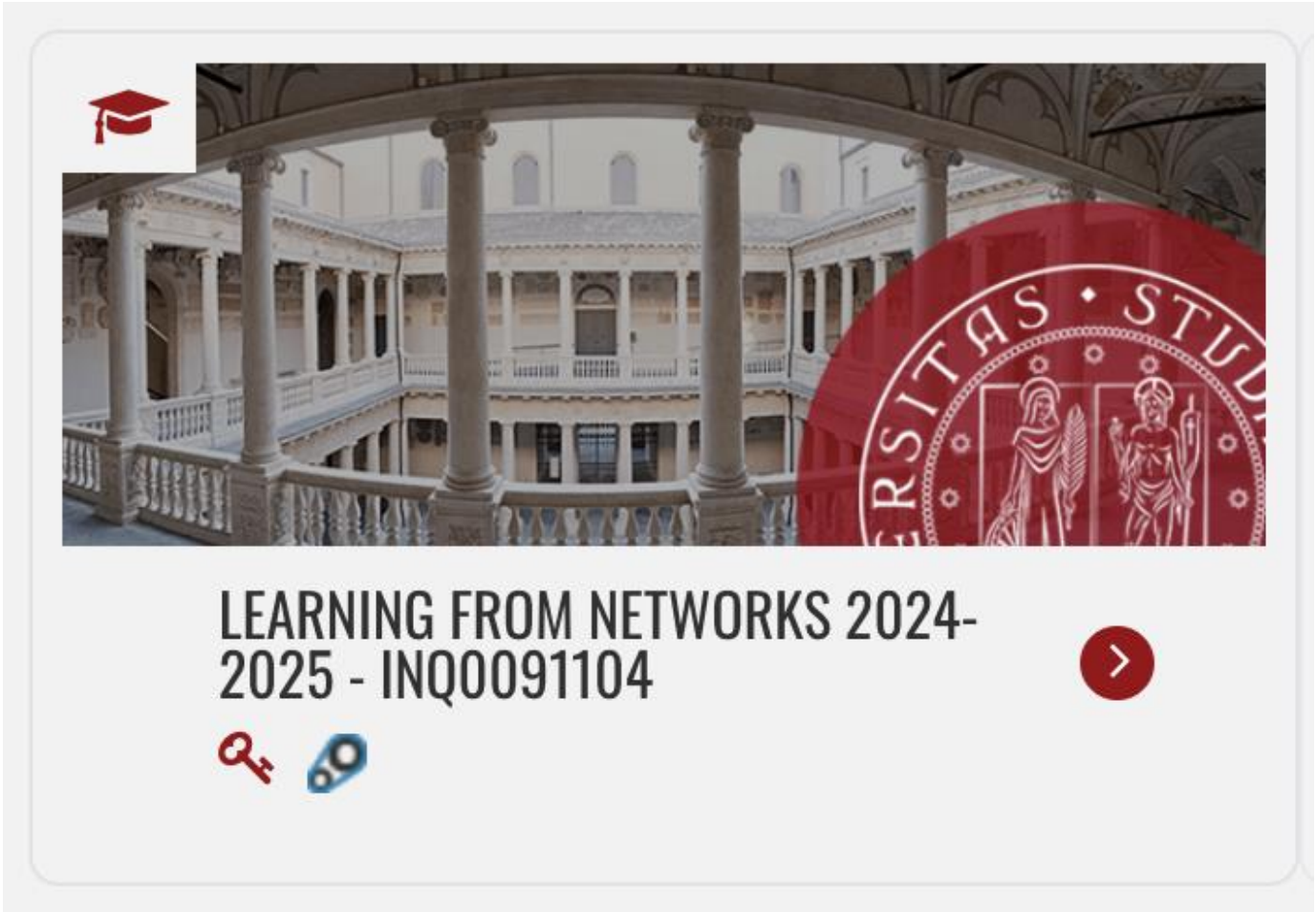
6 credits:

- 48 hours in class lectures
  - Some hours will be in a lab: more decisions to come later
- **102 hours individual study**

Everything (lectures, exams, etc.) in English!

Course Website

Course website: from https://stem.elearning.unipd.it/



Register today if not done yet!

Lectures: When and Where

Wednesday 2:30pm-4:30pm, room Le:

- 2:30pm-3:15pm; 10 mins break; 3:25pm-4:10pm
- 2:30pm-3:15pm; 15 mins break; 3:30pm-4:15pm
- 2:30pm-4:00pm; no break ☹

Thursday 2:30pm-4:30pm, room Ee:

- 2:30pm-3:15pm; 10 mins break; 3:25pm-4:10pm
- 2:30pm-3:15pm; 15 mins break; 3:30pm-4:15pm
- 2:30pm-4:00pm; no break ☹

This is a dynamic course!

Several details may be tuned during the course… check for news and messages from stem.elearning in your email!

Grading

Written test: will be graded on a scale from 0 to 32.

Final project [not compulsory]: max 8 points

Final grade: rounding of written text + final project grades (at least 30.5 for *30 with honors*)

Example

21.5 written test + 6.66 project = 28 final grade

Note: there may be an oral exam just to confirm the grade of the written exam.

Project

There is a **non-compulsory** project (= you can still get the grade of *30 with honors* without doing the project)

Rules:

- groups of 3(?) students
- you choose the topic/problem/data
- deadline to complete the project: January 16th, 2025 (1 week before the first exam)
- max 8 points for the final grade

Important: if you have a ***great project*** you want to work on ***by yourself***, you need my approval within the first 3 weeks of the course.

Project (2)

Three typical types of projects:

- **application project**: pick an application/dataset that interests you, and explore how best to apply what we learn to analyze the dataset
- **algorithmic project**: pick a problem or family of problems, and develop a new algorithm, or a novel variant of an existing algorithm, to solve it.
- **theoretical project**: prove some interesting/non-trivial properties of a new or an existing algorithm

Some projects will combine elements of applications, algorithms and theory.

IMPORTANT: you can also pick a topic related to the analysis of networks not covered by the course!

Project (3)

Useful resources on the course website:

- **page "Datasets and repositories"**: links to datasets or collections of datasets, can be used for the project
- **page "Some interesting papers"**: can provide ideas on motivation, experimental analyses, algorithms, theoretical results, etc.

Feel free to explore other resources!

Project (4)

Deadlines (tentative):

- **project proposal: deadline November 14th, 2025 (7 weeks)**
  - max 2 pages + references with: title, motivation (question, data), method (computational problem, algorithms,…), intended experiments (implementation, machine for experiments,...)
- **midterm report: deadline December 12th, 2025 (11 weeks)**
  - max 2 pages+references: what has changed from the project proposal?
- **final report: deadline January 16th, 2025**
  - max 6 pages + references with detailed description of what has been done + references + links to code (e.g., bitbucket/github repository)

Required for proposal and reports: additional 1 page with detailed contribution of each member, including fraction of the work done by each member **[NOT INCLUDED? 0 POINTS]**

IMPORTANT: If there are issue with the group during the project development, you should let me know before the course ends

Example: somebody does not reply emails and submit a report as if the other 2 people did not do anything

Are you going to do the project?

Final Exam: dates

1. **Wednesday, January 22nd, 2025**
   - time: 9:15am
   - room: Ae

2. **Wednesday, February 5th, 2025**
   - time: 9:30am
   - room: Ke

3. **Tuesday, July 8th, 2025**
   - time: 9:30am
   - room: Ve

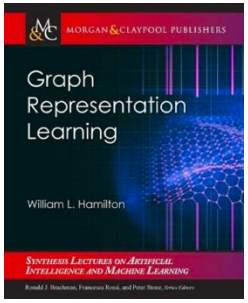
4. **Tuesday, September 9th, 2025**
   - time: 9:30am
   - room: Le

Duration: (usually) 2 hours

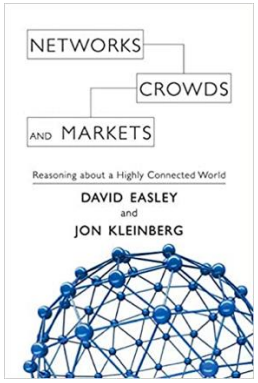
Material

Mostly notes and slides from the lectures

Books for Reference:



- Hamilton, William L, Graph Representation Learning. San Rafael: Morgan & Claypool Publishers, 2020 [**NOTE**: pre-publication openly available version from the author: https://www.cs.mcgill.ca/~wlh/grl_book/]
- Easley, David; Kleinberg, Jon, Networks, Crowds, and Markets. Cambridge: Cambridge University Press, 2010. [**NOTE**: complete pre-publication draft from the authors: https://www.cs.cornell.edu/home/kleinber/networks-book/networks-book.pdf]



Other book:

- Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2009.

Some material will be from the research papers where it has been introduced (will be available on the elearning website)

Material (2)

Additional material: course website (https://stem.elearning.unipd.it/)

- draft slides: provided sometime (…) before the lecture
- slides used in class: published after lectures
- links, etc.

If you are missing background notions (probability, basic graphs notions, etc.): see "Useful links and other stuff" on Moodle and/or ask me for material

Lecturer

Fabio Vandin, Professor, DEI (Department of Information Engineering)

Email: fabio.vandin@unipd.it

Website: www.dei.unipd.it/~vandinfa

Office: 410 (4th floor, DEI/G), phone: 049-827-7946

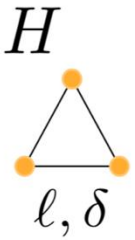
Office hours:

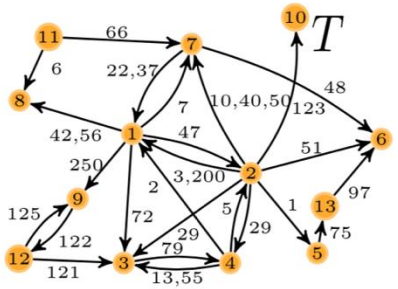
- Monday, 1:15-2:15pm, by appointment (email before Sunday at 5pm)
- Thursday, 4:30-5:15pm, by appointment (email before Wednesday at 5pm)

Lecturer

CV

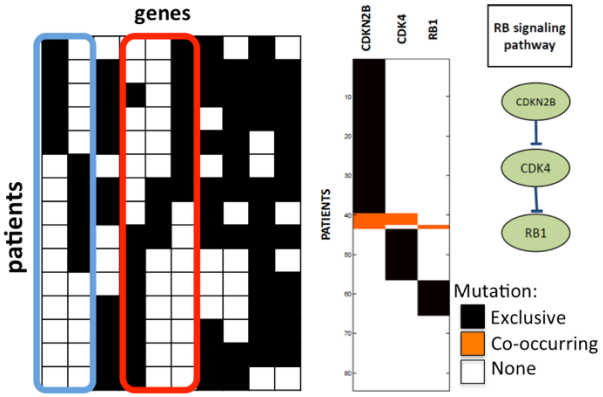
- Laurea Triennale in Computer Engineering (2004)
- Laurea Specialistica in Computer Engineering (2006)
- Ph.D. at DEI (2010)
- 2010-May/2015: Researcher/Assistant Prof.
  - Brown University (USA)
  - University of Southern Denmark
- 2015-2020: Associate Prof. at DEI
- February 2020-now: Prof. at DEI



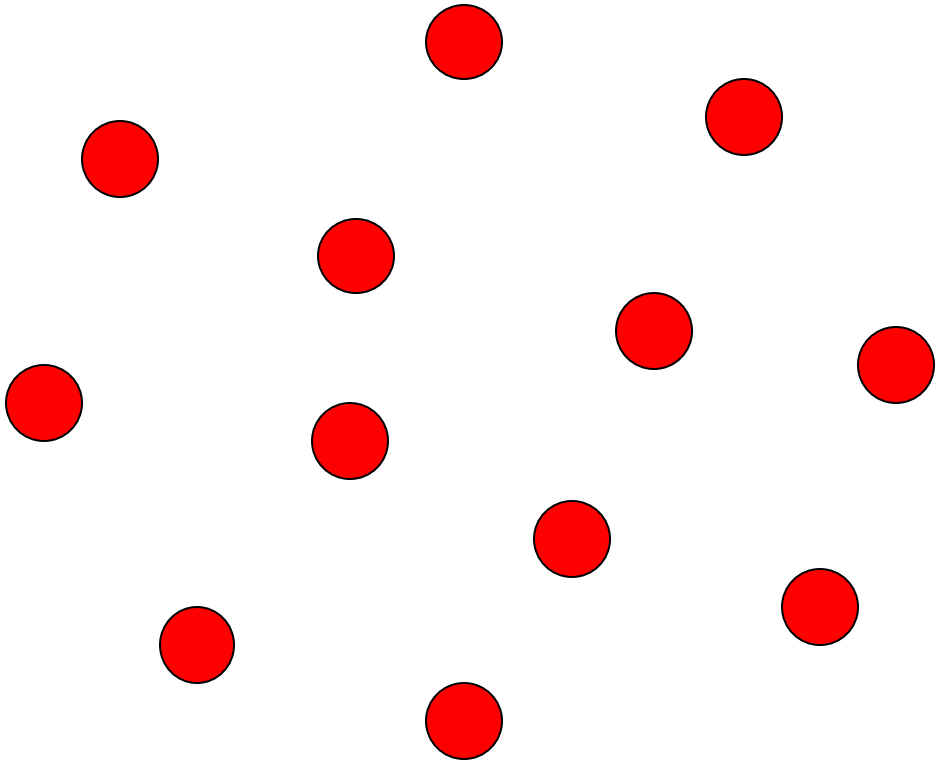


Research Interests:

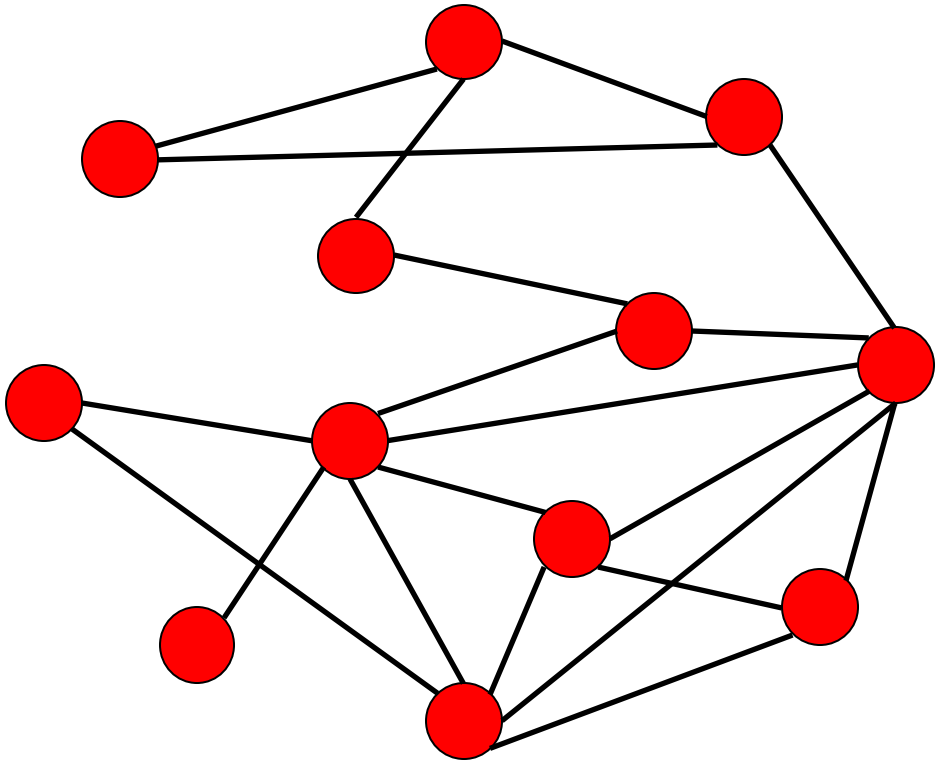
- Methods: algorithms for machine learning, data mining, and big data
- Applications: Biology, Medicine, Social Networks, …



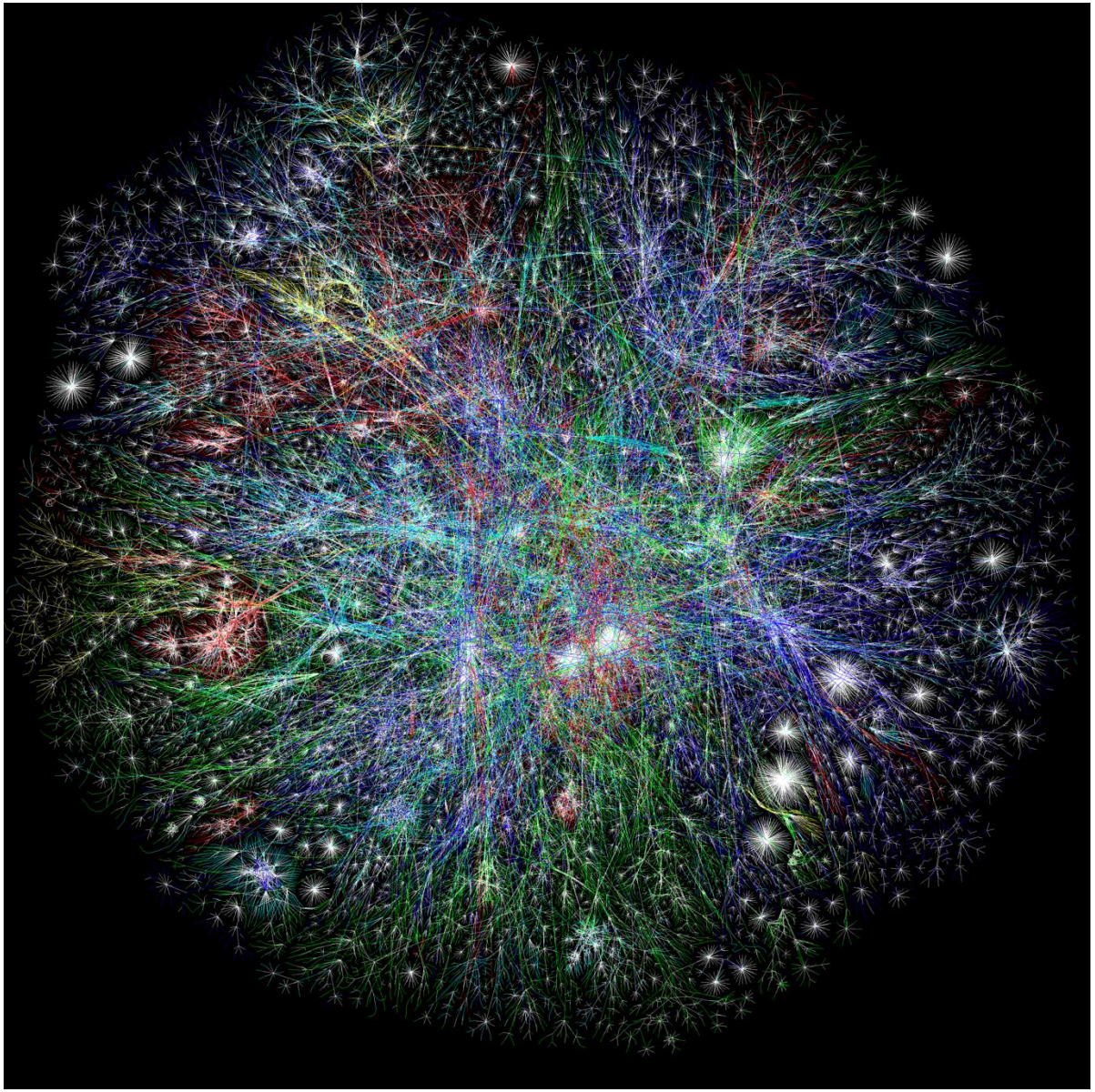
What is a Network?



What is a Network?



What is a Network?



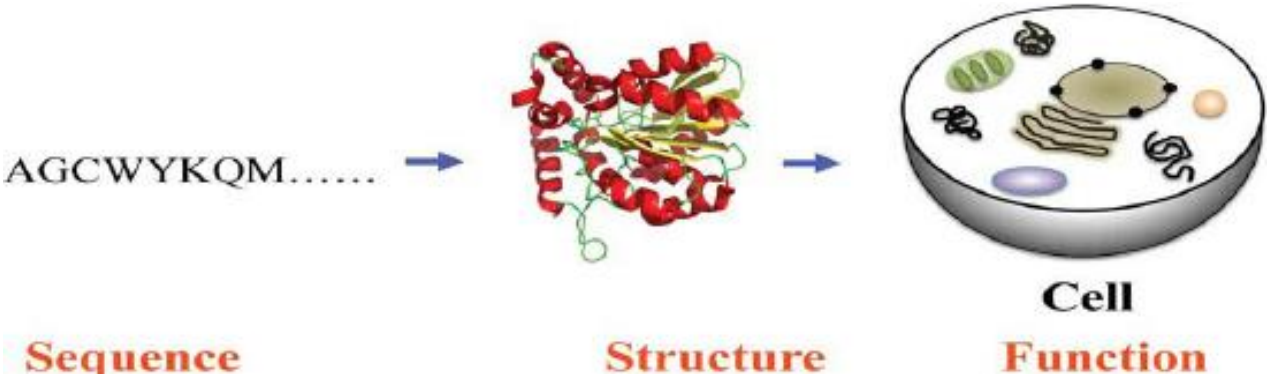
Why Networks?

The relations among objects crucial information that may not be revealed by the object themselves!

Networks appear in every domain!

Example: Protein Function Prediction

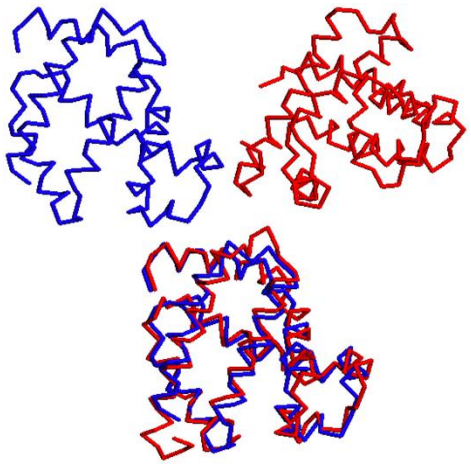
The human cell contains > 80,000 proteins.



Not all of them have a known function.

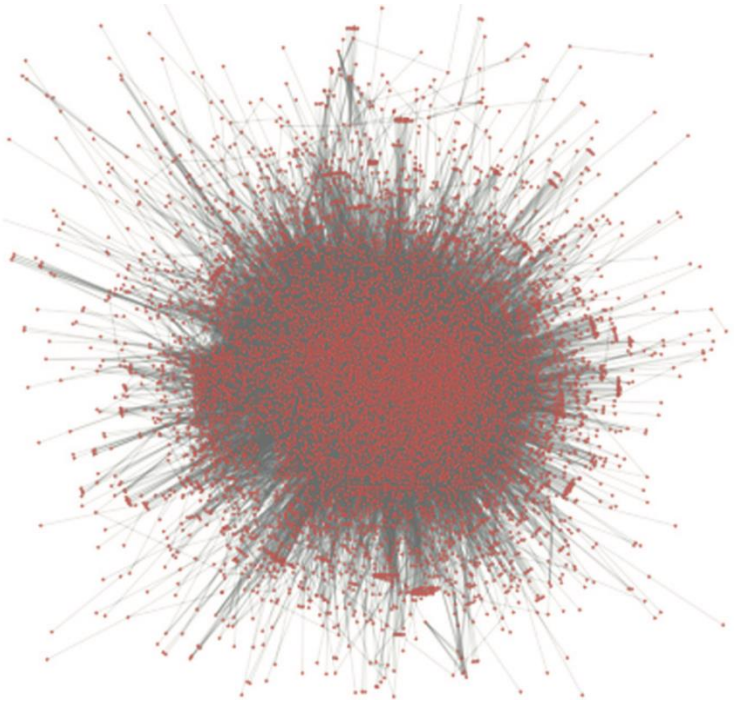
Several approaches to predict a protein function, based on sequence comparison, 3d structures, etc.

```
DISWFSHWGLTSVDMGAPGTAILSTTP-----GESYASYSGTSMATPHVA 392
TMSSFSQWGLASVDMGAPGSAILSTVP-----GGGYSSFSGTSMATPHVA 392
GLAGFSNYGRRTVHVGAPGNSILSTVR-----NNGYGYMSGTSMATPHVA 375
DLAYFSNYG-DWVDLAAPGRAILSTVP-----GGGYASFSGTSMATPHVV 349
NLASFSCYGKKTVHLAAPGVDIYSTVPVT-VFGQGYKHFSGTSMATPYVA 426
NKAGFSHYGSTTVDLGAPGVNIRSTLP-----GNSYGYNSGTSMAAPYVS 430
SLSINSFYSNKYCQLAAPGTNIYSTAPH-----NSYRKLNGTSMAAPHVA 613
DISINSFYSDIYCQVSAPGANIYSTASR-----GSYMEXSGTSMAAPHVA 489
ELDRSSNYGIQRVQVACPGMWILSTVPTSGSSQQPYAEKSGTSMAAPALS 497
SLAESSNYGTKSVQIALPGTDIYSTIPVQERPDDPYGWKTGTSMAAPALS 491
TLSPDSCYSSNYVHLAAPGDDIISTFPQ-----NKYAISSGSSFSAAVVT 382
     * :.    .:. **  * **          *    .*:*::: :
```

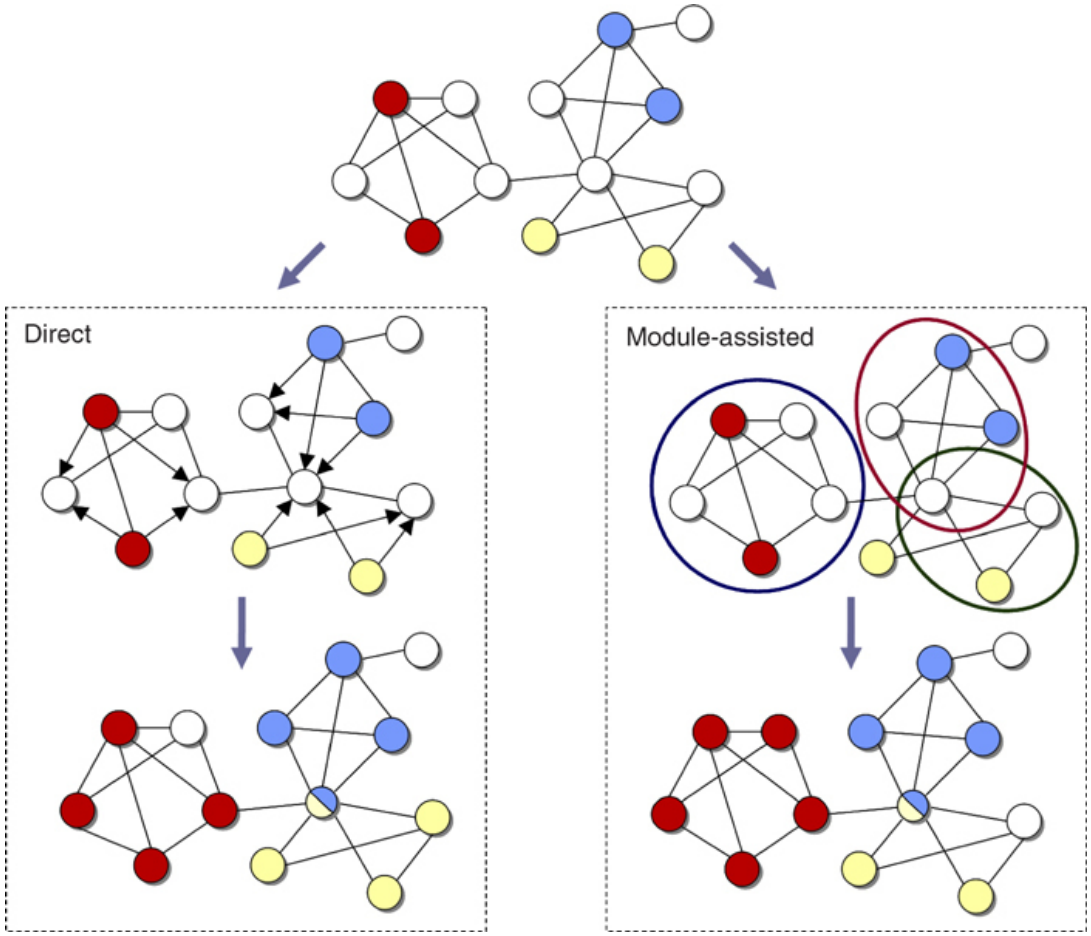


Example: Protein Function Prediction

Successful approach: use protein-protein interaction(PPI) networks to predict the function of uncharacterized proteins



Human PPI network

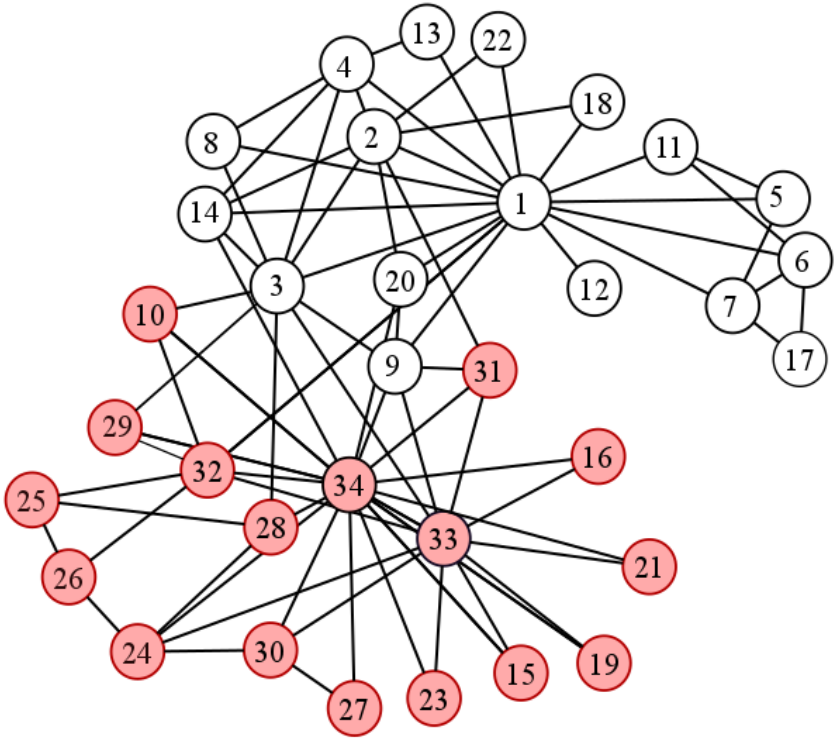


[Sharan et al. Molecular Systems Biology, 2007]

Example: Finding Communities in Social Networks

Studying or predicting social behaviour using networks?

Zachary's karate club network: study of social ties in a karate club (~1970)
Network: interactions outside the club

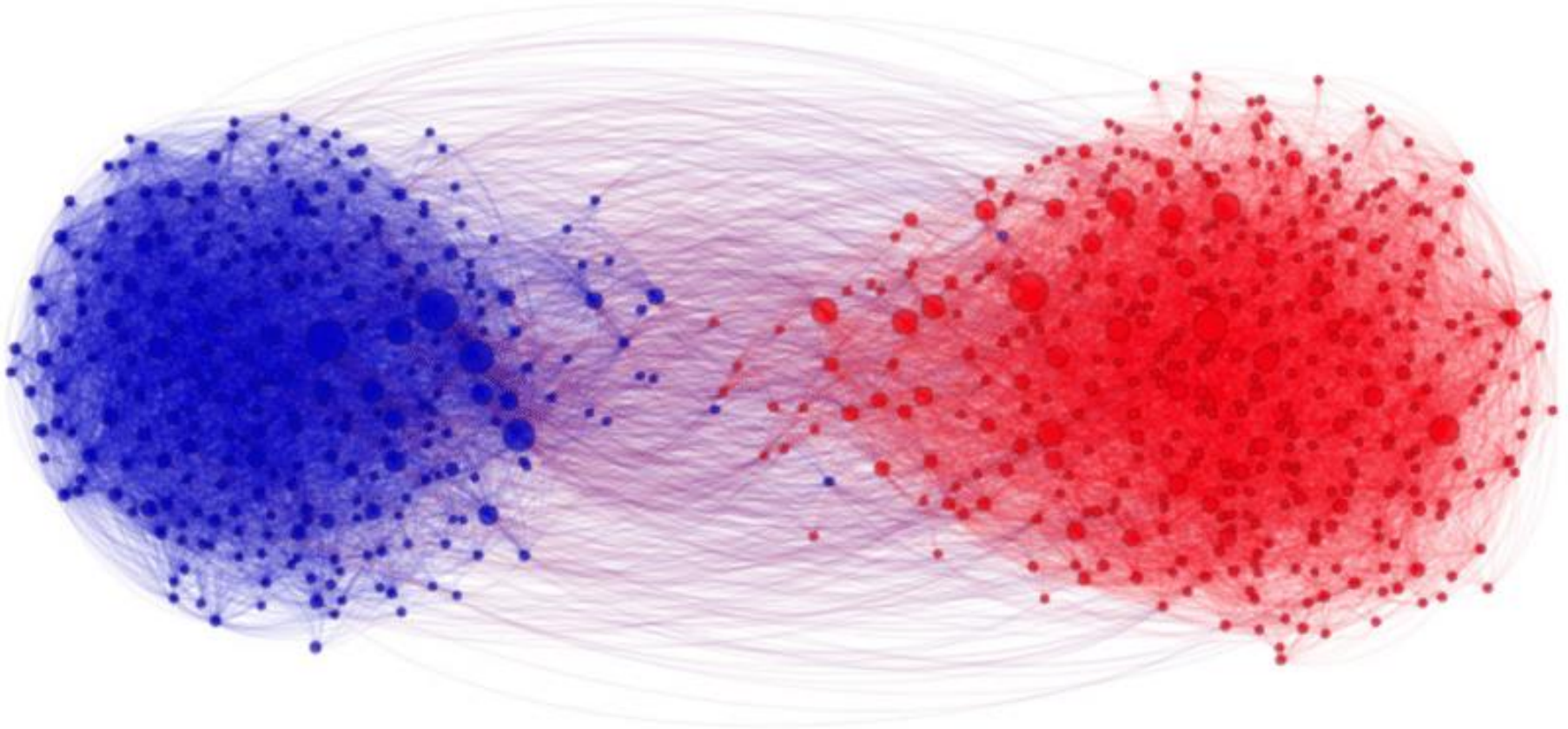


During the study, the club split into 2 clubs

The network could be used to almost perfectly predict the split

[Girvan, M.; Newman, M. E. J. PNAS, 2002]

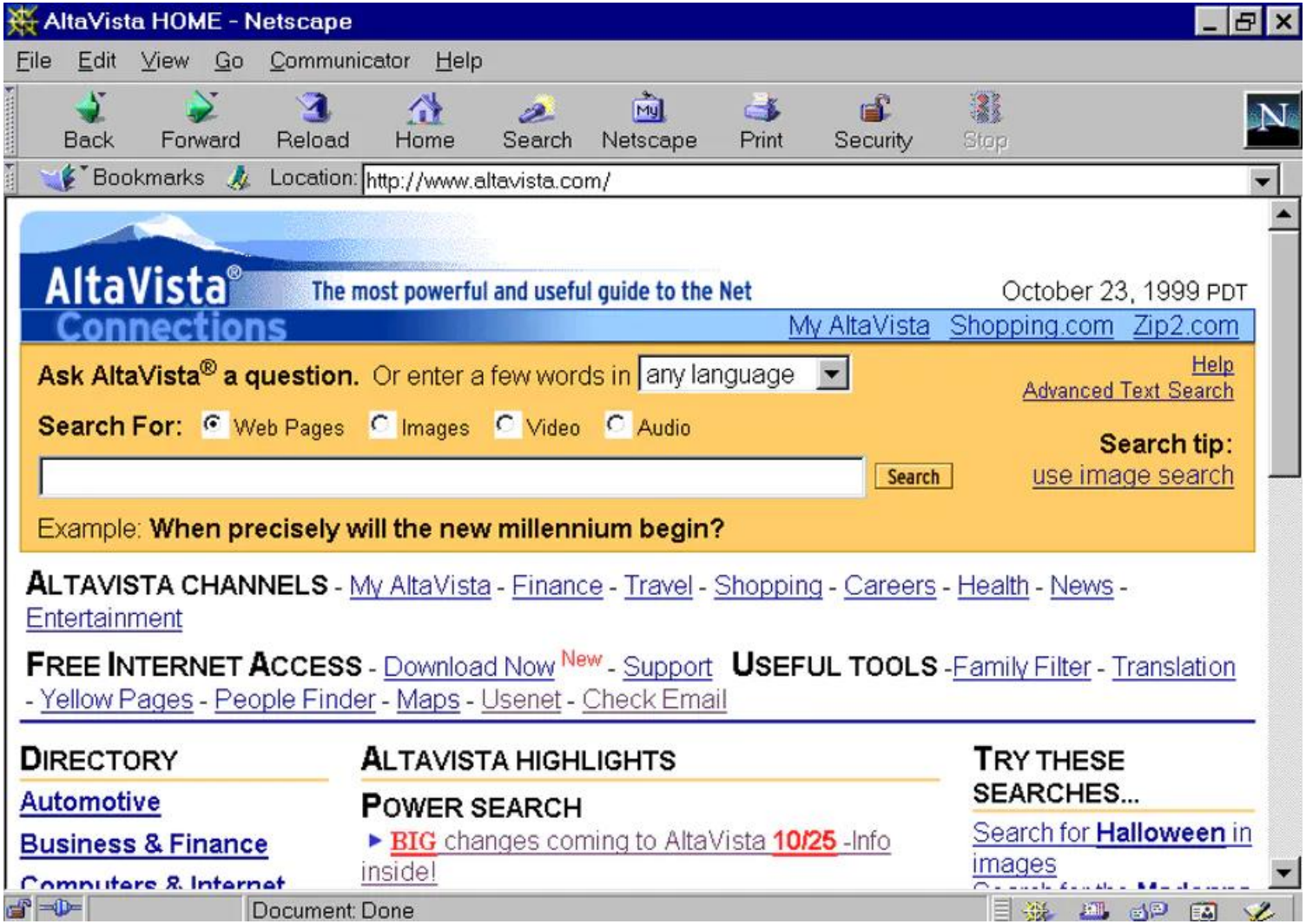
Example: Finding Communities in Social Networks



Network of political blogs during the 2004 U.S. presidential election. Democratic and republican blogs are represented by blue and red circles, respectively

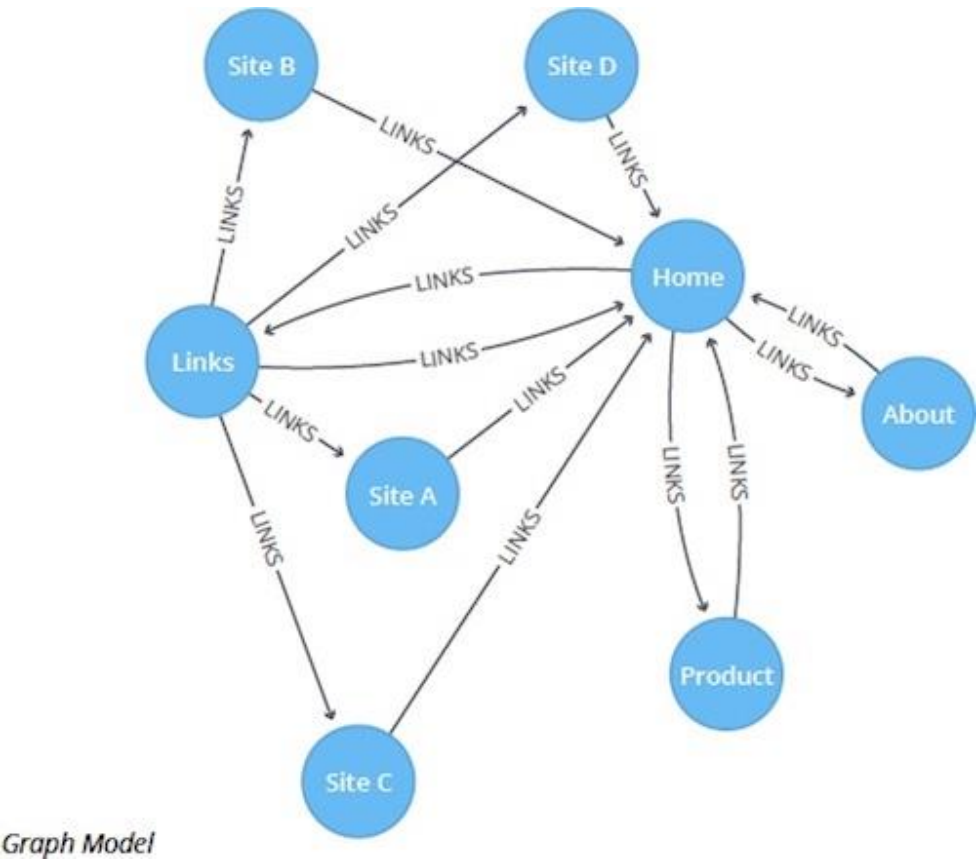
Example: Google's PageRank

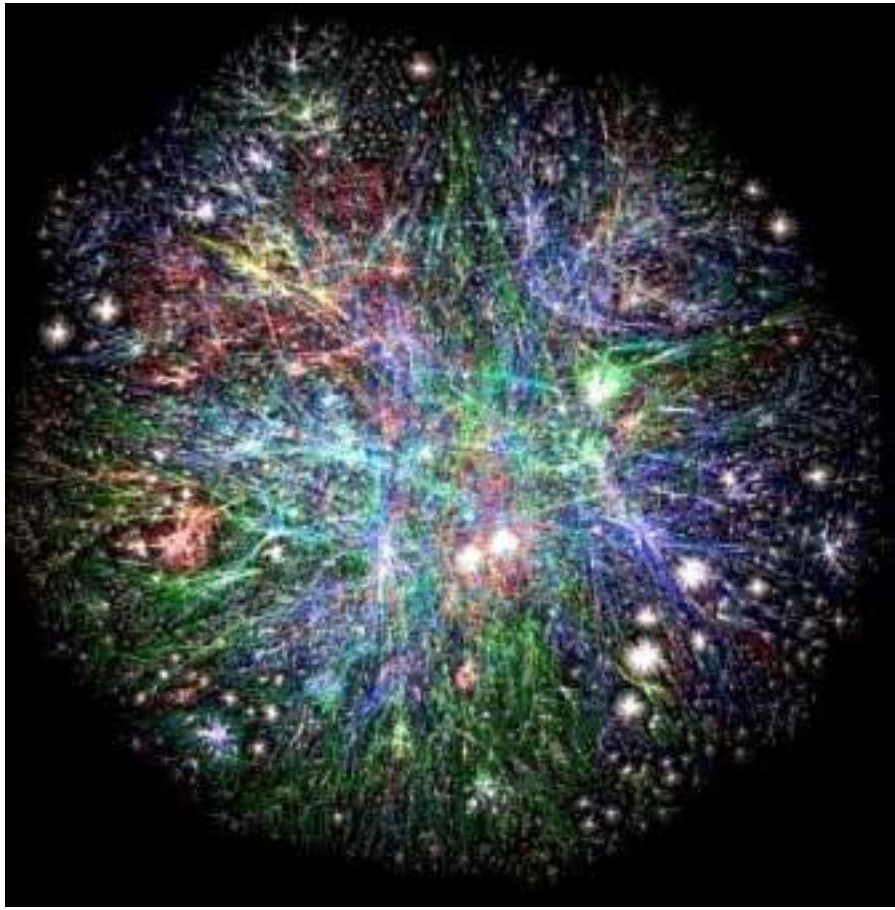
Before Google



Google's idea: what about using the information in the webgraph to rank the results?

Example: Google's PageRank





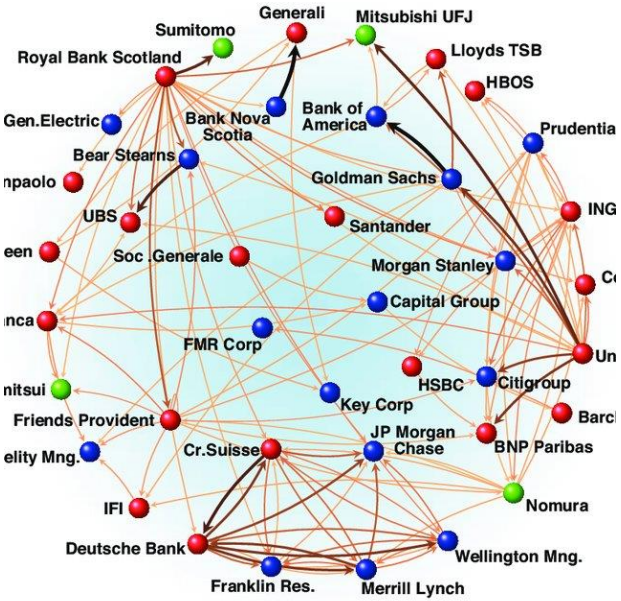
The webgraph

Pagerank centrality ≈ how important a page is in the webgraph

[Page, Brin, et al. Stanford InfoLab, 1999]

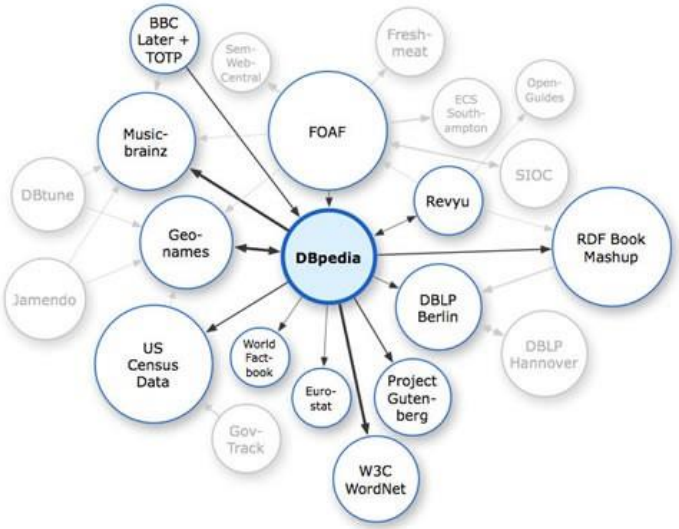
Other Networks

Financial networks

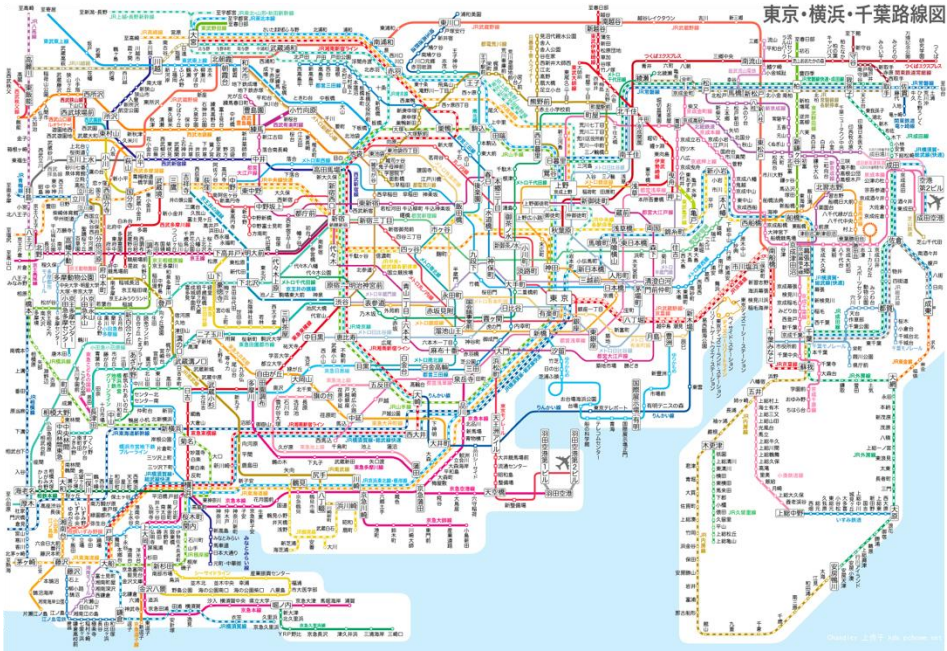


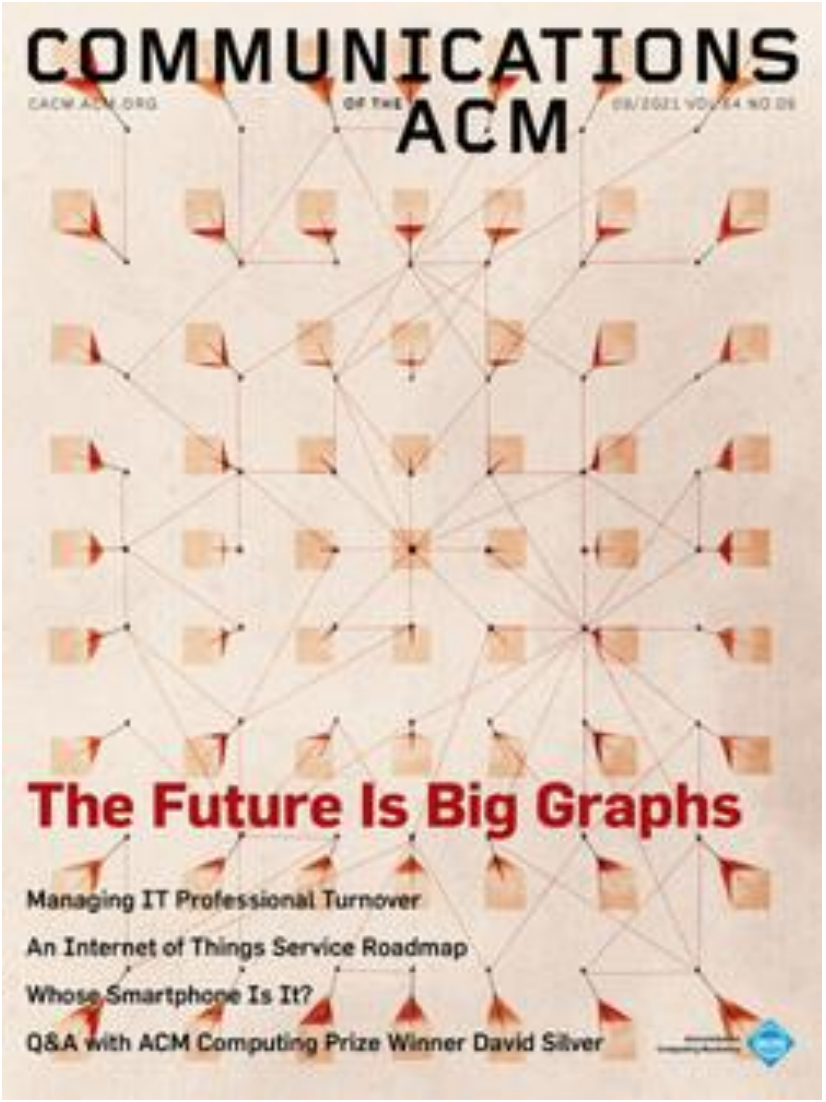
Internet of Things

Information Networks



Underground Networks





“How can future big graph processing and database systems provide highly scalable, efficient, and diversified querying and analytical capabilities, as demanded by real-world requirements?”

Course Goals

Provide the tools to answer the questions:

- how do you learn something useful from network data?
- what type of analyses can you perform?
- what algorithms can we use?

Introduce some of the **basic computational problems** and **tasks** in learning from networks

Present the **most common algorithms** for such tasks

Focus is on the **computation**/**algorithms** and related issues (e.g., scalability)

Course Topics

Introduction and basic notions/algorithms (refresher)

Graph analytics and network features (e.g., centralities, clustering coefficient)

Network patterns and motifs

Network embeddings

Graph neural networks

Network clustering

Advanced topics (e.g., temporal networks, uncertain graphs, polarization bubbles…)

Questions?

Background on graphs

Graph representations

Graph traversal (BFS, DFS) ?

Minimum spanning tree: problem and algorithm?

Single-source shortest path: problem and algorithm?

All-pairs shortest path: problem and algorithm?

Background on probability

Discrete random variables (r.v.)?

Expectation?

Variance?

Independent random variables?

Bernoulli r.v.?

Binomial r.v.?