T-61.6040
Seminar course on information networks

Lecture 1: Introduction

12.09.07

Tervetuola!

- Who we are
  - My name is Gemma Garriga
  - Assistant is Nikolaj Tatti
- Language: English
- You can contact us at the lectures, or in the office (third floor) by appointment
- Course webpage

  http://www.cis.hut.fi/Opinnot/T-61.6040/
- For contacting

  t616040@cis.hut.fi

About this course

- This is a postgraduate course (lisenssiaattikurssi)
- If you pass, you get 6 credit points
- Timetable: Wed 14-16 in Lecture Hall T4
- Calendar: From 12.09.07 to 12.12.07

About this course

- This is a postgraduate course (lisenssiaattikurssi)
- If you pass, you get 6 credit points
- Timetable: Wed 14-16 in Lecture Hall T4
- Calendar: From 12.09.07 to 12.12.07

Attention: No lecture on 19.09.07!

About this course

- Goals of the course
  - To read and discuss about interesting papers on information networks
  - Understand the properties and techniques
  - Get a global view on everything and more specific view (through homework and presentation) on at least three topics of your choice
- Prerequisites
  - Background on algorithms, graph theory, probabilites and linear algebra
- Style
  - Seminar style
  - Reading list at

    http://www.cis.hut.fi/Opinnot/T-61.6040/

Topic list

Papers in the reading list are organized by topic sections:

- Network models
- Power laws and scale free networks
- Small-Worlds: Search and properties
- The Web Graph
- Web search, Link analysis, Spectral analysis
- Propagation effects in networks: gossips, epidemics and trust
- Clustering and community structure
- Games and networks
- Biological networks

Topic list

Papers in the reading list are organized by topic sections:

- Network models
- Power laws and scale free networks
- Small-Worlds: Search and properties
- The Web Graph
- Web search, Link analysis, Spectral analysis
- Propagation effects in networks: gossips, epidemics and trust
- Clustering and community structure
- Games and networks
- Biological networks

- Links to other similar seminar with other topics
  - ▷ P2P networks ...

Requirements to pass the course ...

To pass the course you have to:

- Participate in the lectures

  Only one absense is allowed. For a good grade, it is important to generate discussion, ask questions to the speaker ...

- Prepare a presentation

  Choose one paper from one of the topics and prepare a **45 min** presentation

- Homework

  Two homework assignments. Each assignment consists on preparing a reaction paper

- Final project

  A more practical task

Requirements to pass the course ...

Details of the presentation

- To do before next Wed:
  - Send an email with your preferred paper for the presentation
  - Send an email to inform about a couple of preferred slots from all the Wednesdays between 26.09.07 to 12.12.07
- Papers will be handed out in first come first serve fashion. Presentations will be scheduled within constraints
- Note that some papers are quite long, then you can choose to present a subpart of the paper. Presentations should not be too simple nor too overloaded

Requirements to pass the course ...

Details of the homework

Preparing a reaction paper means:

- Read at least two closely related papers relevant to one of the sections of the course. Chosen papers for the assignment should be:
  - From a topic different from the one you presented
  - Not presented in the class by anyone
- You should then write approx 3 pages addressing the points:
  - Summary of the technical contents of the papers
  - Discussion of why the papers are interesting in relation to the chosen section of the course
  - Discussion of weaknesses and how they could be improved
  - Discussion of strenghts and promising research questions that arise from the papers

Requirements to pass the course ...

Details of the project

- Still not available. Will be posted in the web page of the course.

Deadlines

- 24.10.07 First homework assignment
- 21.11.07 Second homework assignment

- 19.12.07 Final project

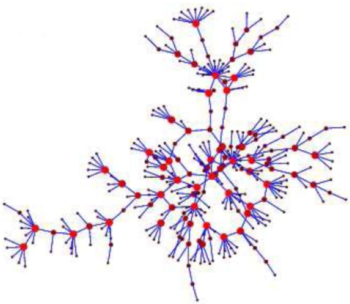
Web page

- All details in the course web page

  http://www.cis.hut.fi/Opinnot/T-61.6040/

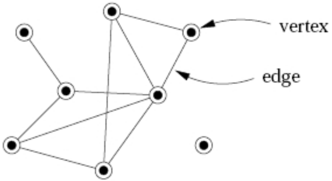
- Questions about the course organization?

Lecture 1: Introduction to networks



Information networks

- A network is a set of entities (vertices or nodes) with connections between them (edges or links).



- An edge between two nodes reflects an interaction.
- This interaction shows an information exchange, hence we have information networks.
- Networks are everywhere

Networks in the real world

Networks arise from different branches of science:

- Social networks
- Knowledge networks
- Technological networks
- Biological networks

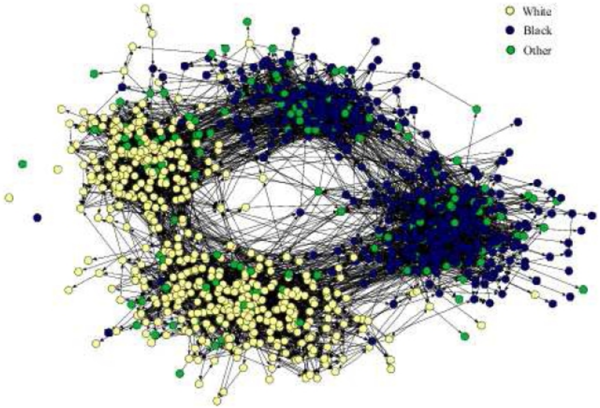
Networks in the real world

Networks arise from different branches of science:

- **Social networks**
- Knowledge networks
- Technological networks
- Biological networks

Social networks

- Links denote a social interaction
  - E.g. friendship network of children in a US school



Social networks

- Other social networks:
  - Actor networks
  - Co-authorship networks
  - Co-appearence networks
  - Telephone call networks
  - Email communication networks
  - Bluetooth networks
  - ...

Social networks

An important set of experiments on social networks are the famous small-world experiments of Milgram in 1967.

- The experiments probed the distribution of path lengths in an acquaintance network
- Experiment: pass a letter to one of your first-name acquaintances in an attempt to get it to an assigned target individual
- Result: Most of the letters were lost, but about a quarter reached the target and passed on average through the hands of only 6 people
- This experiment coined the popular six degrees of separation concept

Social networks

Social networks are called asortative

- Let $\mathcal{K}_{nn}(k)$ be the average degree of the nearest neighbors of the vertices of degree k
- In social networks, the value of $\mathcal{K}_{nn}(k)$ is an increasing function of k
- In contrast, knowledge, technological and biological networks are typically dissasortative.

Networks in the real world

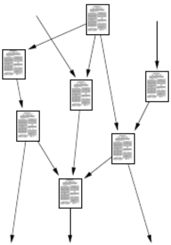
Networks arise from different branches of science:

- Social networks
- **Knowledge networks**
- Technological networks
- Biological networks

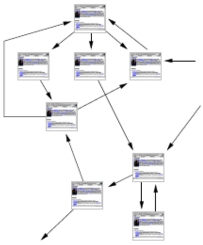
Knowledge networks

Nodes store information. The two best studied knowledge networks are:

- Citation network (directed acyclic)
- The Web (directed)



citation network



World–Wide Web

Knowledge networks

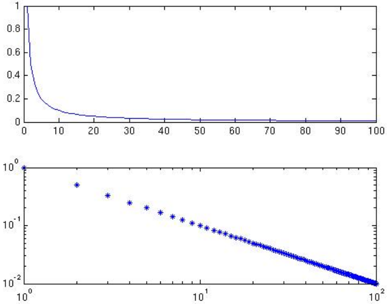
Large citation networks started to become available around 1960's.

- Law of scientific productivity
  - The distribution of the numbers of papers written by individual scientists follows a power law
  - That is, the number of scientists who have written k papers falls off as $k^{-\gamma}$, for some constant γ
  - Indeed, both in-degree and out-degree of the network follow a power law

Knowledge networks

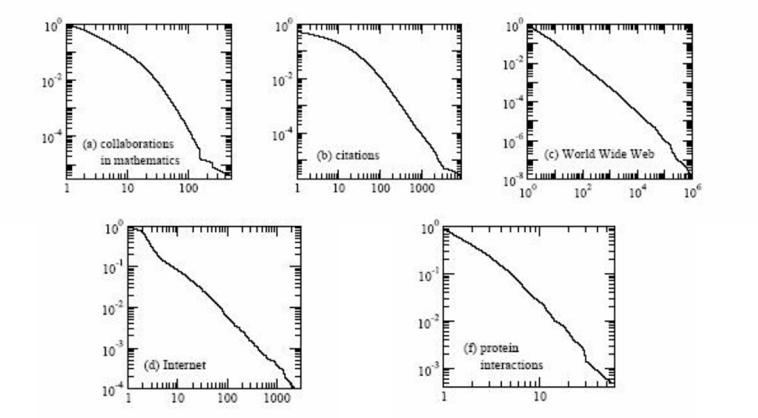
Power law signature

- f_k = fraction of nodes with degree k
  = probability of a randomly selected node to have degree k
- Power law means $f_k = Ck^{-\gamma}$
- Right-skewed/Heavy-tail distribution
- Power law distribution gives a line in the log-log plot, that is, $\log f_k = -\gamma \log k + \log C$



Examples

Power law signature



Knowledge networks

World Wide Web is the network of webpages containing information, linked together by hyperlinks.

- Not to be confused with the Internet
- The Web also appears to follow a power law of in-degree and out-degree
- The picture of the Web is biased though, as it comes from crawls of the network. E.g. low in-degree might be an underestimate

Knowledge networks

Other knowledge networks:

- Peer to peer networks
- Word networks
- Networks of trust
- . . .

Networks in the real world

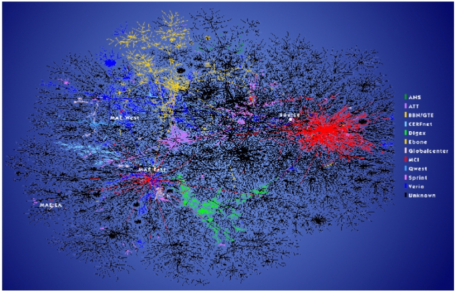
Networks arise from different branches of science:

- Social networks
- Knowledge networks
- **Technological networks**
- Biological networks

Technological networks

Man-made networks designed for distribution of commodity or resource.

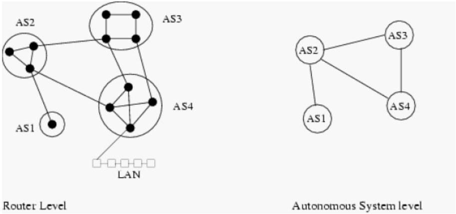
- Electric power grid network
- Networks of roads, railway, pedestrian traffic
- Natural networks, such as rivers
- Mail networks
- The Internet, i.e. network of physical connections between computers



Technological networks

The Internet:

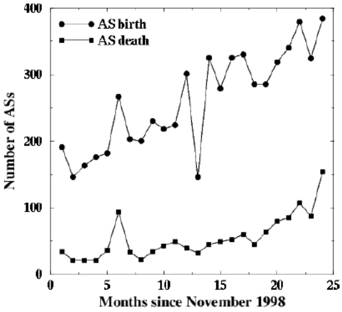
- For a reasonable representation, first we have to individuate the basic units of Internet
- Two levels of resolution:
  - ▷ Router level (IR)
  - ▷ Autonomous system level (AS)
- Connections between ASs correspond to the aggregation of traffic among their respective routers



Technological networks

Nature of the growth at the AS level:

- Keep track of the monthly number of new AS that appear in the maps, and the number of AS that disappear (birth and death of AS)

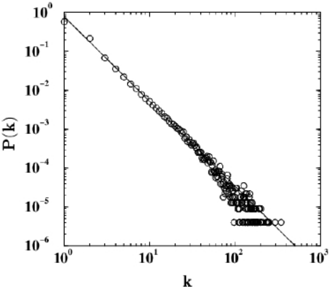


(Quian et al, 2002)

- The Internet growth is not driven by the simple addition of new AS and connections, but is the result of a complex birth-death process

Technological networks

- Both IR and AS level maps exhibit a degree distribution decaying as a power law, $k^{-\gamma}$ with $\gamma \simeq 2.1$



- Internet is a so-called scale-free network

Networks in the real world

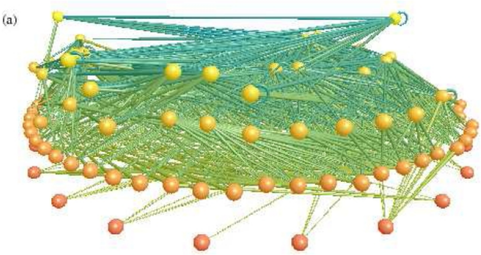
Networks arise from different branches of science:

- Social networks
- Knowledge networks
- Technological networks
- **Biological networks**

Biological networks

Biological systems represented as networks

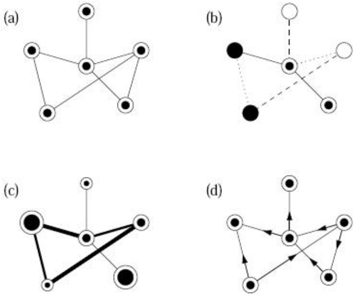
- Network of metabolic pathways
- Gene regulatory networks
- Food web networks
- Vascular networks
- Protein-protein interaction networks
- ...



What comes next?

Difficulties in studying these kind of systems are

- ▷ Heterogeneous nature of entities and interactions



- ▷ Size and scale problems
- ▷ Complex pattern of the interactions between entities

Network properties

What can we study?

- The small world effect
- Degree distributions
- Network resilience
- Patterns, gossips, trusts, ...
- Search properties and algorithms
- Community structure
- Network navigation
- Evolution of statistics: degree correlation, diameter, transitivity coefficient ...
- . . .

Topic list

Papers in the reading list are organized by topic sections:

- Network models
- Power laws and scale free networks
- Small-Worlds: Search and properties
- The Web Graph
- Web search, Link analysis, Spectral analysis
- Propagation effects in networks: gossips, epidemics and trust
- Clustering and community structure
- Games and networks
- Biological networks
- Links to other similar seminar with other topics

References

Partially based on:

- M. E. J. Newman, The structure and function of complex networks, SIAM Reviews, 45(2): 167-256, 2003.
- M. Newman, D. Watts, A.-L. Barabási, The Structure and Dynamics of Networks (Princeton University Press, 2006).