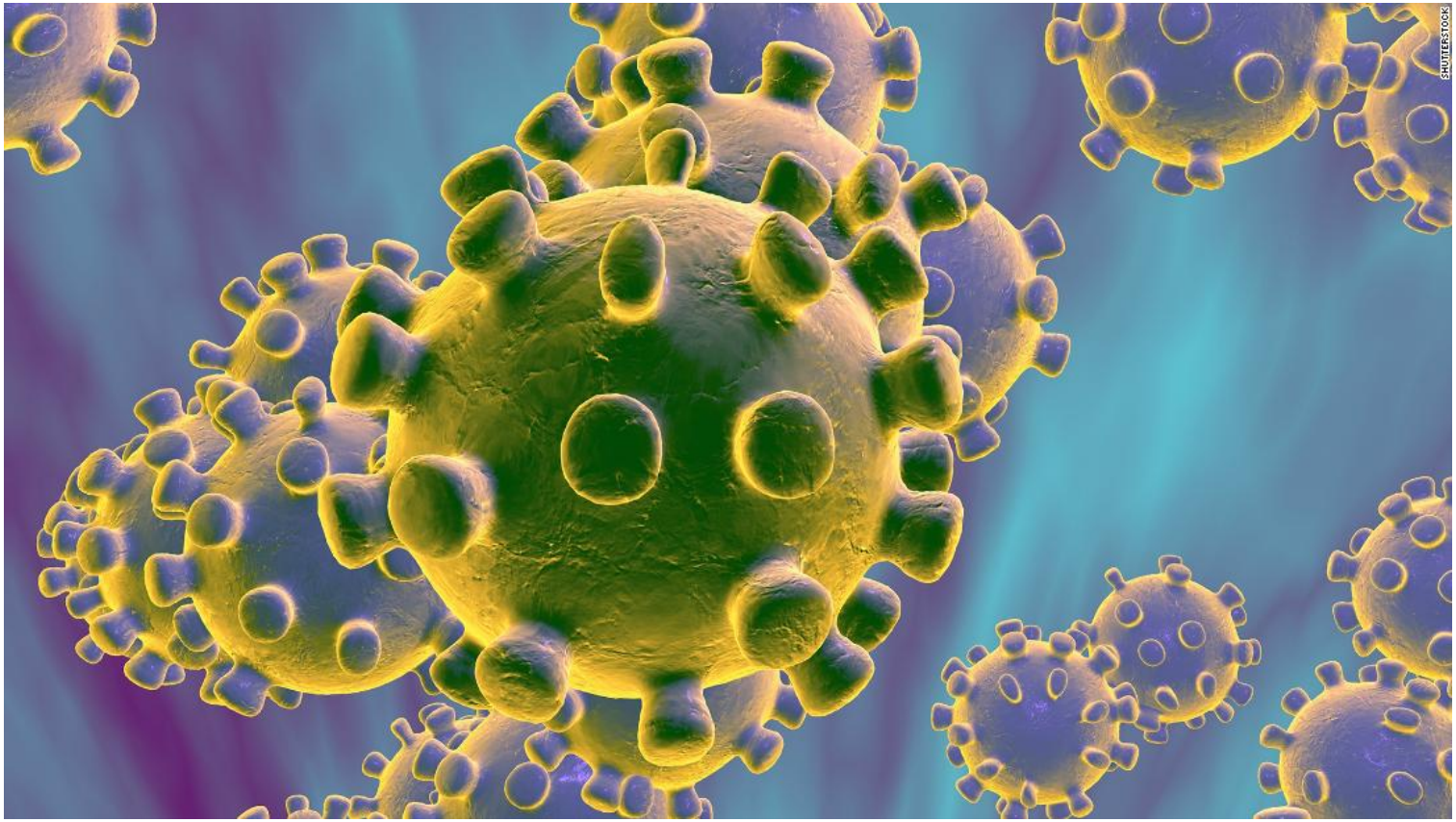
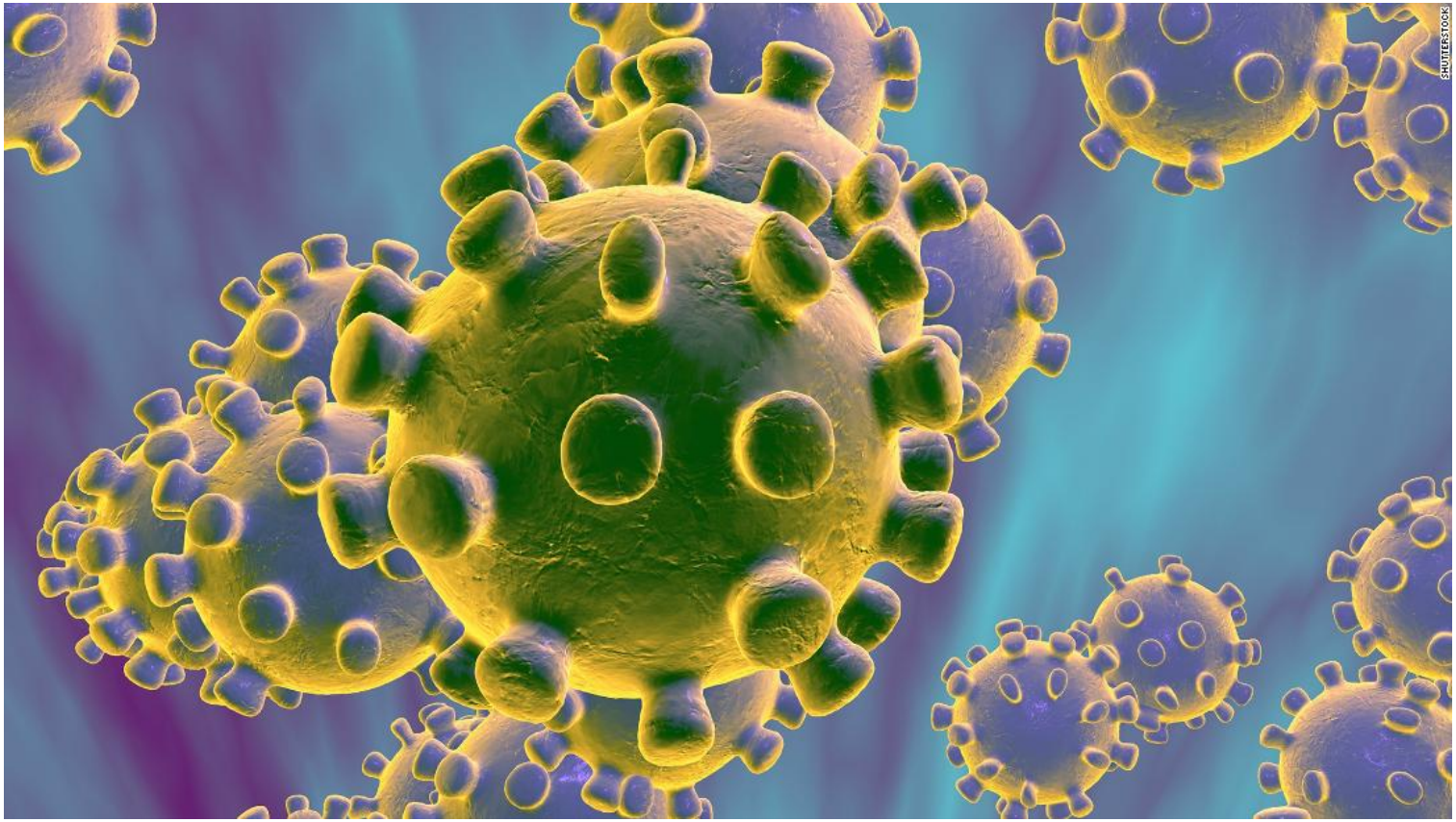
Computational Social Systems 1 (VU) (706.616)

Elisabeth Lex

ISDS, TU Graz

March 5, 2020


SHUTTERSTOCK



Corona virus: epidemic

JoanneWrightForCongress @JWrightforCA34 · Feb 28

Doesn't @BillGates finance research at the Wuhan lab where the Corona virus was being created? Isn't @georgesoros a good friend of Gates? Isn't it always when @HillaryClinton tweets that fire and brimstone hits us? Check Gates Foundation and Clinton Foundation for stock sells.

> **Esther** @esther_speaks · Feb 28
>
> I forgot about Soros; the ugly monster always looking for the opportunity to produce a tragedy in America But he's going to be exposed also, maybe that's the only way we're going to get rid of this chief evil spirit. twitter.com/reubing/status...

308 151 223

Show this thread

[1] https://news.un.org/en/story/2020/02/1056672

Corona virus: infodemic[1]

[1] https://news.un.org/en/story/2020/02/1056672

Book tip: Christakis, N. A., & Fowler, J. H. (2011). Connected: The surprising power of our social networks and how they shape our lives. Little, Brown Spark.

Modeling such processes

- Spread of virus & spread of (mis-)information: dynamical processes
- Modeling epidemics in (social) networks topic in this class!

Lecturer

Name:	Elisabeth Lex
Office:	ISDS, Inffeldgasse 13, 5th Floor, Room 072
Office hours:	By appointment
Phone:	+43-316/873-30841
email:	elisabeth.lex@tugraz.at

Lecturer

Name:	Denis Helic
Office:	ISDS, Petersgasse 116, Room 26
Office hours:	Tuesday from 12 til 13
Phone:	+43-316/873-30610
email:	dhelic@tugraz.at

Language

- Lectures in English
- Communication in German/English
- If in German: please informally (Du)!
- Examination: German/English

Course context

- Computational Social Systems 1 (VU) (706.616)
- Replacement and further development of the previous course Web Science

- Elective course in Bachelor or Master Computer Science/Software Development and Business

Goals of the course

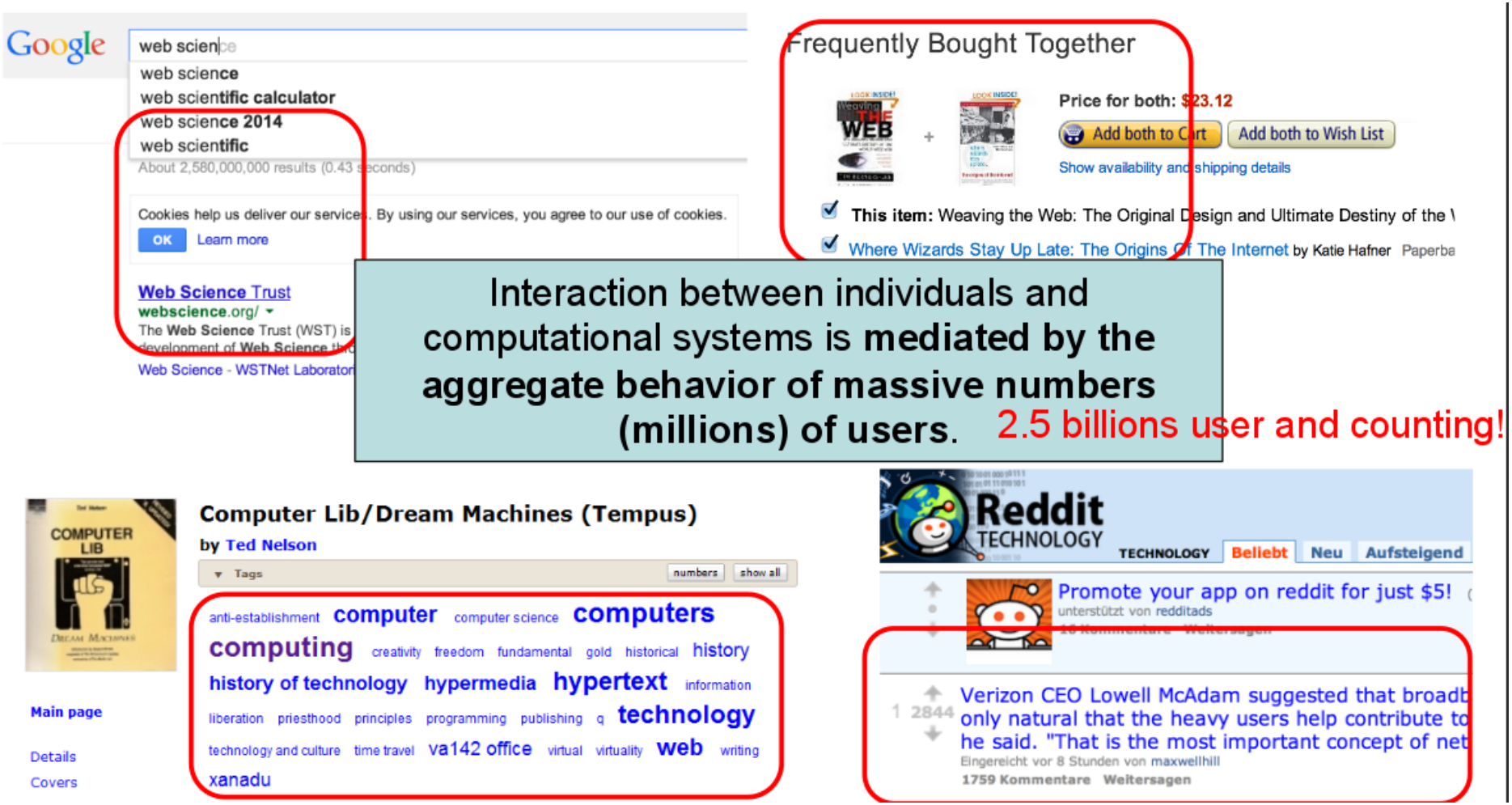
(1) Learn about basic scientific methodology for computational social systems (CSS): network theory and analysis, data mining, social processes on the Web

(2) Understand that further development of the CSS requires scientific engineering approach

- Science: analyze Web as object of scientific inquiry and learn something new
- Engineering: use new knowledge and improve algorithms

(3) Learn about python as analysis tool

Computational Social Systems



Computational Social Systems

- Computational social systems (CSS): algorithms and tools enable & shape social behavior
- Dynamic interaction between algorithms, tools, and users shape outputs of CSS and thus, again user behavior
- Examples: the Web, online social networks, collaboration platforms, social media

Computational Social Systems

What does that mean?

- Beginning: software engineers like us
- Full control over system, its behavior, its functionality
- But: Algorithms work with data generated by users
- We engineers lost control of the system!

Examples: Algorithms based on user generated data

- Recommender Systems (Amazon, Netflix, ...)
- Ranking algorithms in search engines
- PageRank: links created by users

All can have unintended side effects! Which ones?

Side effects of algorithms on the Web

- Amplification of biases (e.g. gender, ethnicity)
- Distribution of false information (e.g. misinformation, conspiracy theories)
- Boosting of malicious content (e.g. spam farms)

Scientific methodology to study CSS

- Observe: e.g. collect access logs on a recommender site
- Measure: e.g. how many users buy which products
- Quantify: e.g. similarity between users
- Make a model: e.g. users with similar interests buy similar products
- Predict with the model and validate: e.g. implement and evaluate
- Apply: engineering approach to implementing the model in the software

Example of a CSS: The Web

- Fastest growth of any technology in human history
- How long did it take to reach 50 million people?
  - Telephone 75 years
  - Radio 35 years
  - TV 13 years
  - The Web 4 years

The Web

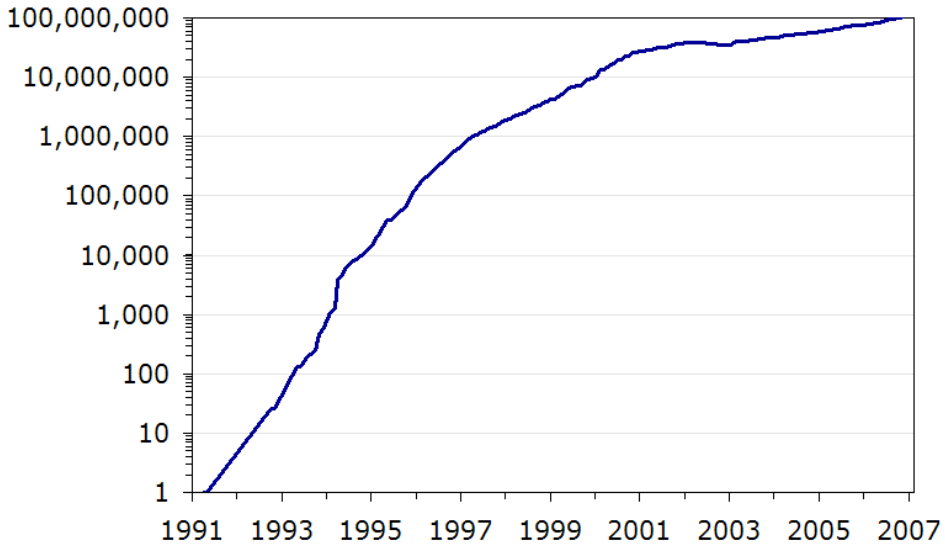


Figure: Jakob Nielsen, 100 Million Web Sites,
http://www.useit.com/alertbox/web-growth.html

Three growth stages:

- 1991-1997: Explosive growth, rate of 850% per year.
- 1998-2001: Rapid growth, rate of 150% per year.
- 2002-2006: Maturing growth, rate of 25% per year.

The Web

The size of the Web

- $\approx$ 1000 billions http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html
- Pages indexed by Google $\approx$ 50 billions http://www.worldwidewebsize.com/

The beginnings of the Web...

- 1989: A scientist from CERN wanted to share information with scientists from CERN and other universities: proposal for World Wide Web by Tim Berners-Lee

The beginnings of the Web...

- 1989: A scientist from CERN wanted to share information with scientists from CERN and other universities: proposal for World Wide Web by Tim Berners-Lee

Hypertext

"HyperText is a way to link and access information of various kinds as a web of nodes in which the user can browse at will. It provides a single user-interface to large classes of information (reports, notes, databases, computer documentation and on-line help). We propose a simple scheme incorporating servers already available at CERN... A program which provides access to the hypertext world we call a browser..."

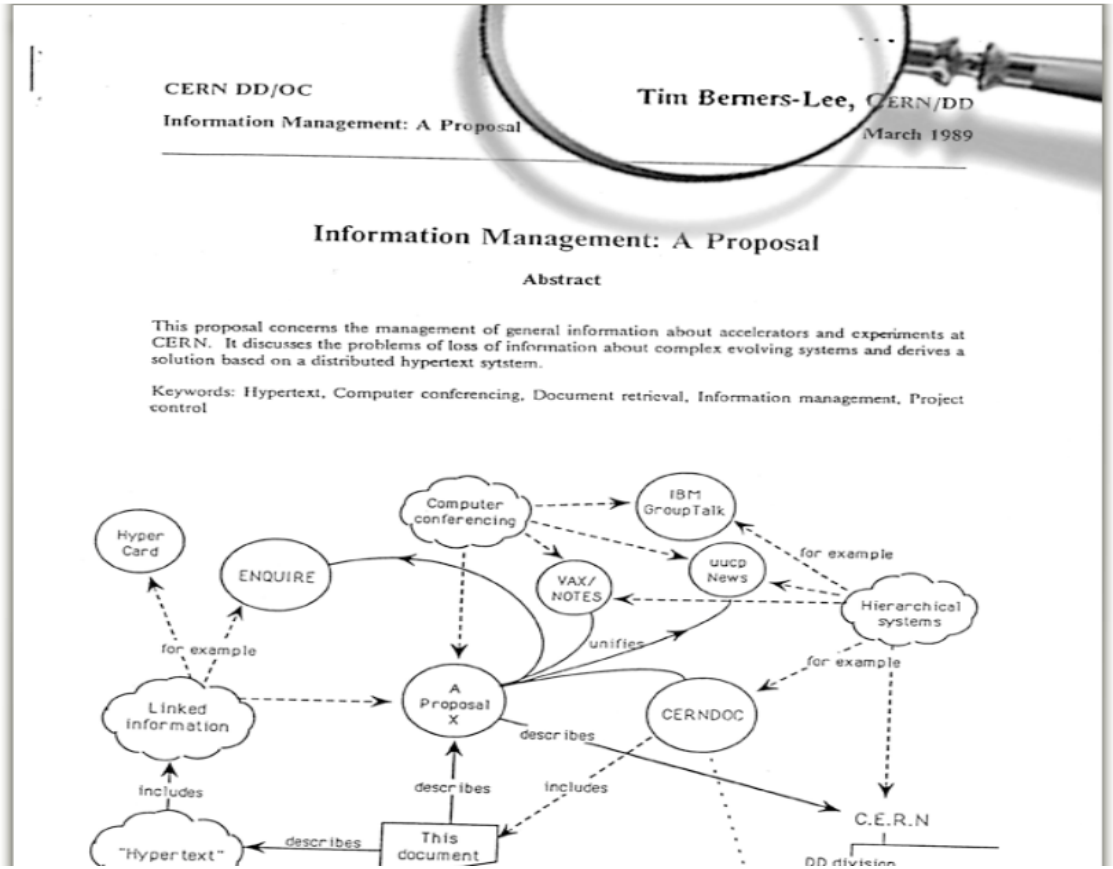
Tim Berners-Lee , R. Cailliau. 12 November 1990, CERN

The beginnings of the Web...

Full paper rejected at HT'91

http://www.informatik.uni-trier.de/~ley/db/conf/ht/ht91.html

Accepted as a poster: "An Architecture for Wide Area Hypertext", Hypertext '91 Poster Abstract, *SIGLINK* Newsletter



CERN DD/OC

Information Management: A Proposal

Tim Berners-Lee, CERN/DD

March 1989

Information Management: A Proposal

Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext sytstem.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control

The beginnings of the Web...

- URL: a unique address of a Web resource (page, image, ...)
- HTTP: a stateless protocol built on the top of TCP/IP for data exchange
- HTML: a simple markup language for Web documents
- <a href=...>: an element for linking other documents on the Web
- Simple, scalable, flexible: reasons for the huge success

Theories on the Web...

A few examples of assertions:

- Every page on the web can be reached by following less than 10 links. (True/False/Depends?)
- A Wikipedia page contains, on average, 0.03 false facts (True/False/Depends?)
- 1%-4% of users express their search queries in the form of goals such as "increase adsense revenue" (True/False/Depends?)
- The average number of words per search query is more than 3 (True/False/Depends?)

Theories on the Web...

- Can these statements be easily validated?
- Can they lead to good/interesting theories?
- What constitutes good theories?
  - Clarity, simplicity
  - Predictive and explanatory power
  - Applicability and utility
  - Testability and falsifiability

Networks

- Network theory
- Graph theory vs. Network theory
- Graph theory: focus on mathematics and analytical approaches (small graphs)
- Network theory: focuses on networks observed in the "real world"
- Large empirical networks, statistical approaches
- Many different forms of networks available on the Web
- Can you name a few of them?

Networks

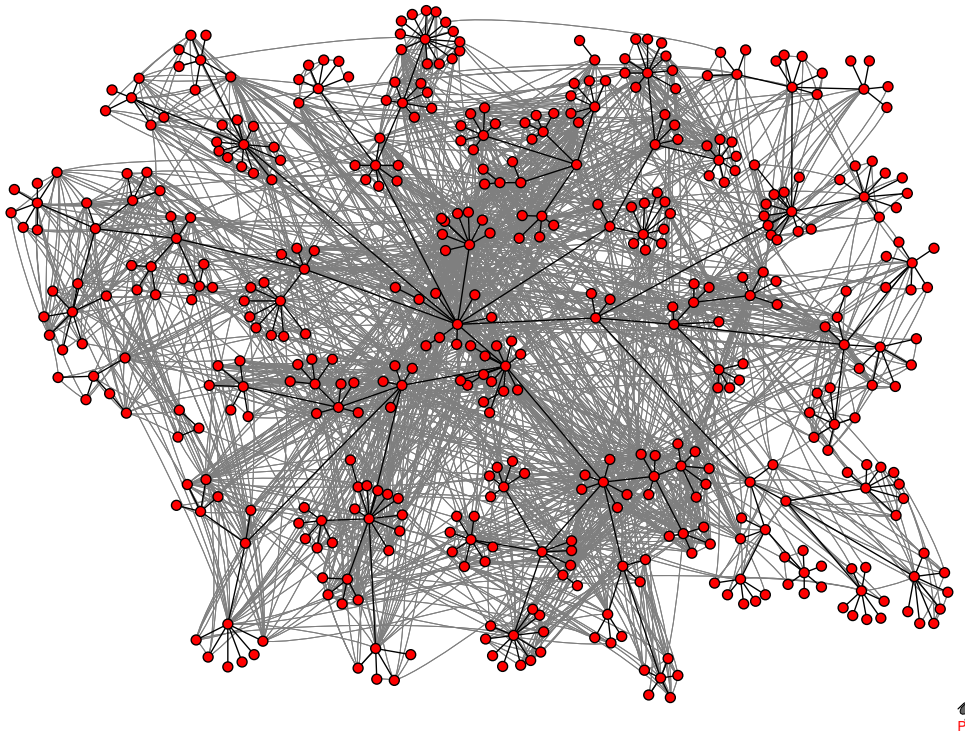


Figure: Social network of HP Labs constructed out of e-mail communication.
From: How to search a social network, Adamic, 2005.

Networks

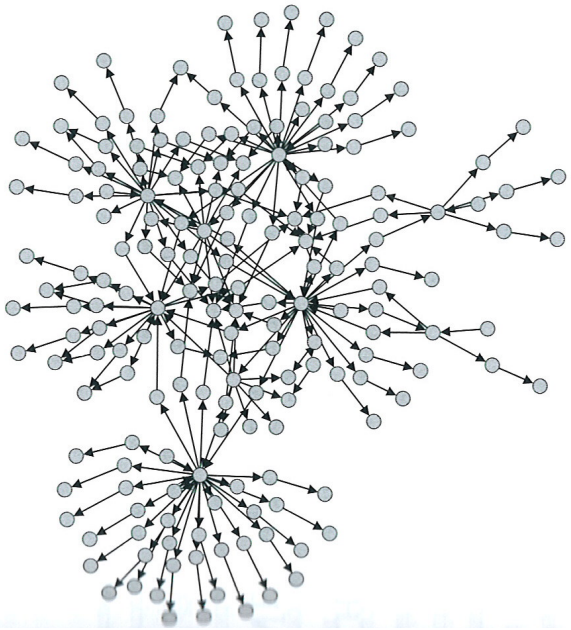


Figure: Network of pages and hyperlinks on a Website. From: Networks, Mark Newman, 2011.

Networks

- Content of this course is mainly based on free, online textbook:
- Networks, Crowds, and Markets: Reasoning About a Highly Connected World, by David Easley and Jon Kleinberg, 2010
- http://www.cs.cornell.edu/home/kleinber/networks-book/

Agents

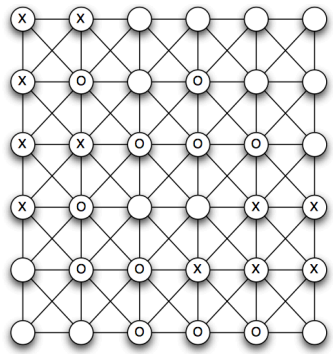
- Agents, interactions, dynamics
- Individual agents follow simple rules
- Interaction guided by simple rules
- However, evolution may result in a very complex system
- Emergent properties

The Schelling Model (1/2)

- Population of agents of type X or O
- Types: immutable characteristics (e.g., age)
- Init: two type of populations placed at random on grid

x	x				
x	o		o		
x	x	o	o	o	
x	o			x	x
	o	o	x	x	x
		o	o	o	

(a) *Agents occupying cells on a grid.*



(b) *Neighbor relations as a graph.*

The Schelling Model (2/2)

- Determine if each agent is satisfied with its current location
- Satisfied if surrounded by at least t of its own type of neighbors
- Agents unsatisfied: Agents move to the next random location
- Threshold t applies to all agents in the model
- Is this realistic?

  No. Why?

The Schelling Model (2/2)

- Determine if each agent is satisfied with its current location
- Satisfied if surrounded by at least t of its own type of neighbors
- Agents unsatisfied: Agents move to the next random location
- Threshold t applies to all agents in the model
- Is this realistic?

  No. Why?

  In reality, agents have different thresholds

Example

X1*	X2*				
X3	O1*		O2		
X4	X5	O3	O4	O5*	
X6*	O6			X7	X8
	O7	O8	X9*	X10	X11
		O9	O10	O11*	

(a) Initial stage

X3	X6	O1	O2		
X4	X5	O3	O4		
	O6	X2	X1	X7	X8
O11	O7	O8	X9	X10	X11
	O5	O9	O10*		

(b) After one round

Figure: Left image: dissatisfied agents have an asterisk. Right image: shows new configuration after all dissatisfied agents have relocated

What is the threshold t?

Effects of Schelling Model

- What can happen if agent relocates?
- Other become unsatisfied
- What can happen in the long run?
- Segregation of population

Course logistics

- Course website:
  https://courses.isds.tugraz.at/dhelic/websci/index.html
- Slides will be made available on course website
- Additional readings, references, links, etc. also on website

Grading

- Two homeworks with applied mathematics/programming problems
- Homeworks solved individually
- Homeworks will be announced in TeachCenter (coming soon)
- A small (scientific) project with a presentation at the end of the term
- Project in groups of two

Grading

- Each homework: 25 points for a total of 50 points
- Project: 30 points
- Total for the project and homeworks is 80 points
- You have to reach at least **41** points combined to be positive!

Grading

- 0-40 points: 5
- 41-50 points: 4
- 51-60 points: 3
- 61-70 points: 2
- 71-80 points: 1

Questions?

- Raise them now (+1 +1)
- Ask after the lecture (+1)
- Visit us in the office hours (+1)
- Send us an e-mail (-1)
- As a side note: you should(!) interrupt me immediately (+1 +1 +1) and ask any question you might have during the lecture

Course Topics

- World Wide Web: a short history
- What is network theory? Why it is relevant for the Web?
- How do networks evolve?
- How can we model dynamics of social influence and aggregating beliefs on the Web?
- How do we search in networks?

How many of you know...

- 6 degrees of separation (Small-world phenomenon)
- Degree distribution
- Power-law networks
- The meaning of PageRank
- Agent-based modeling

Course calendar

- 05.03.2020: Introduction and Motivation
- 12.03.2020: Networks I
- 19.03.2020: Networks II (Random Graphs)
- 26.03.2020: Small World Phenomenon I
- 02.04.2020: Small World Phenomenon II

Course calendar

- 23.04.2020: Power Laws and Preferential Attachment I
- 30.04.2020: Power Laws and Preferential Attachment II
- 07.05.2020: Network Dynamics I (Bayesian Learning, Information Cascades)
- 11.05.2020: Network Dynamics II (Agent-based Modeling)
- 14.05.2020: Network Dynamics III (Opinion Dynamics)

Course calendar

- 28.05.2020: Information Networks I (Hubs and Authorities)
- 04.06.2020: Information Networks II (PageRank)
- 18.06.2020: Project presentations I
- 25.06.2020: Project presentations II

Questions?

- Raise them now (+1 +1)
- Ask after the lecture (+1)
- Visit us in the office hours (+1)
- Send us an e-mail (-1)
- As a side note: you should(!) interrupt us immediately (+1 +1 +1) and ask any question you might have during the lecture