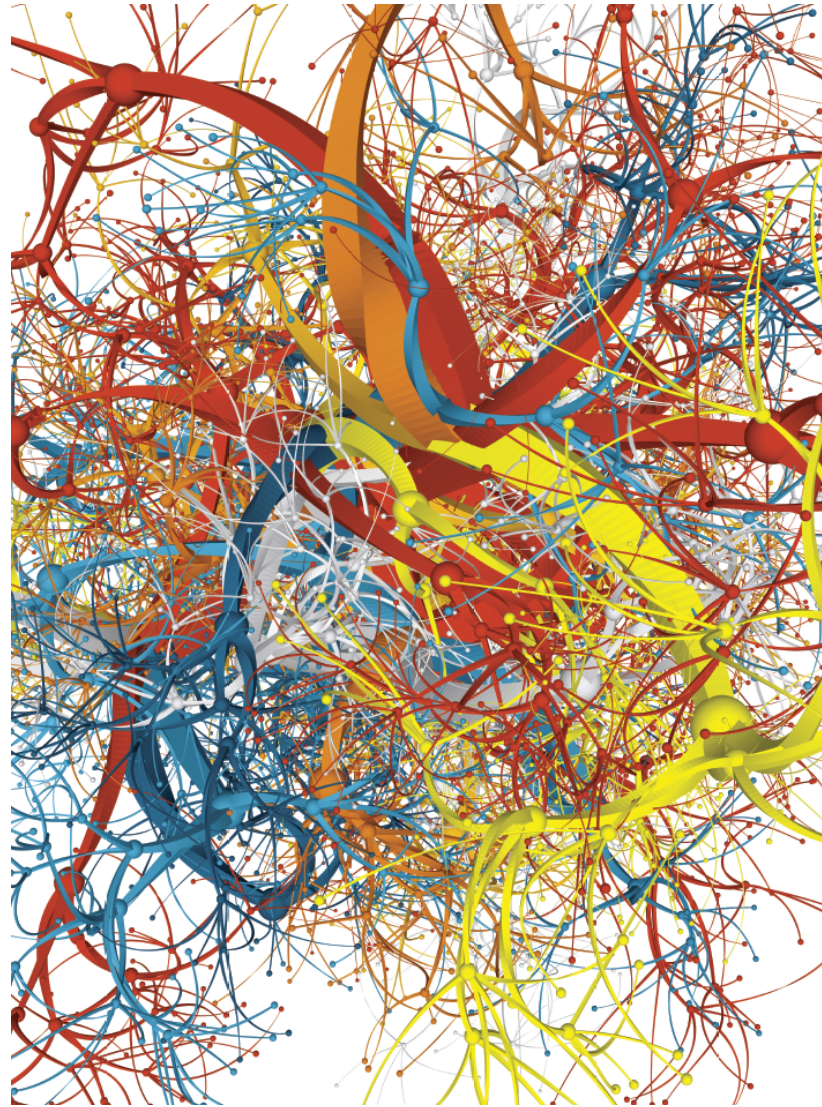


# Web Science



<http://dl.acm.org/citation.cfm?id=1364782.1364798>

# Overview

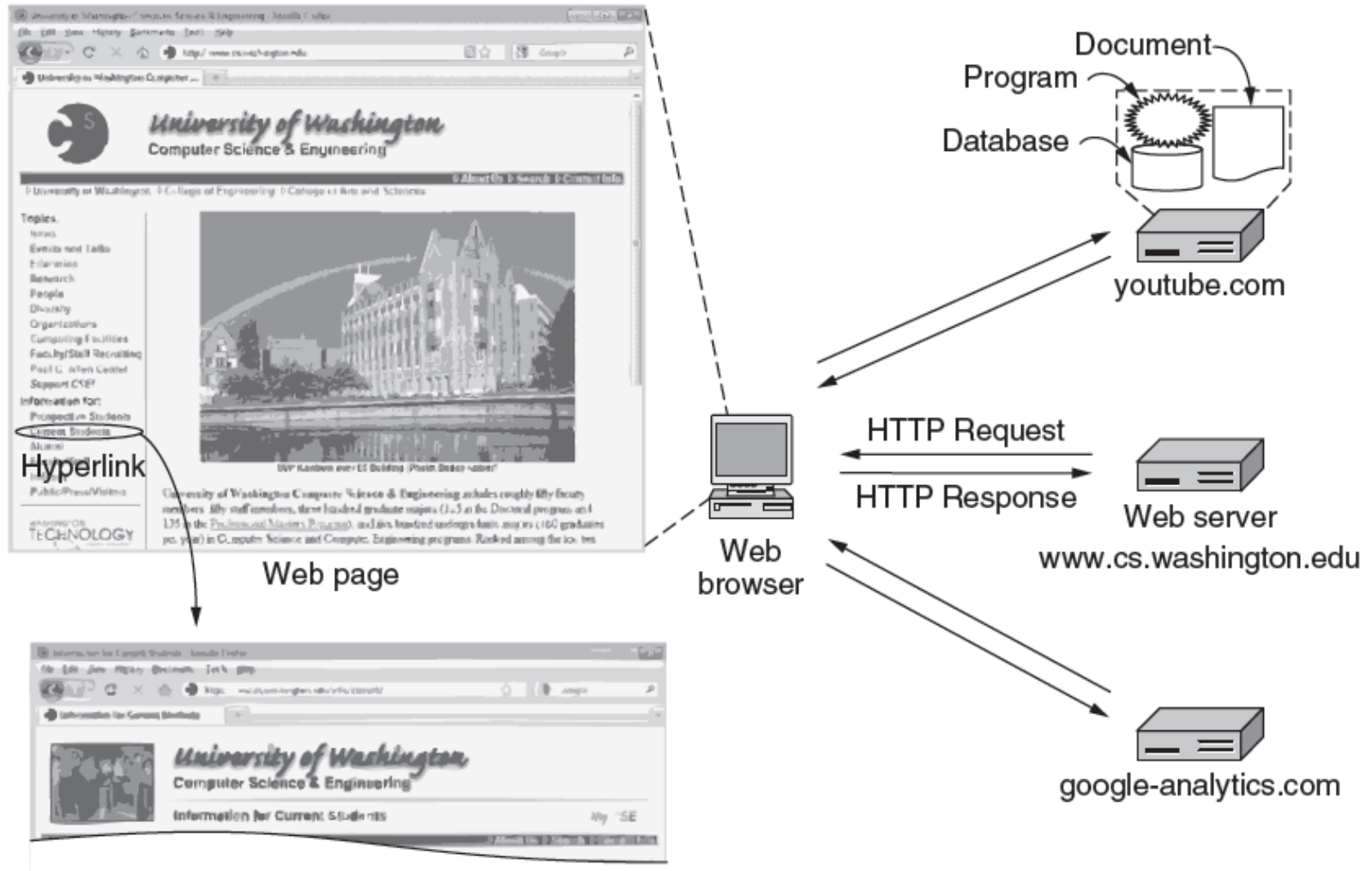
- The web
  - History
  - Stats
- "Web science"
  - History of the term
- What we're going to do
  - Course details
  - Topics

# A short history of the web

- **1989** Tim Berners-Lee at CERN
- **1990** HTTP/0.9, HTML, URLs, first text-based browser
- **1993** Marc Andreessen releases NCSA Mosaic, graphical browser
- **1993** CERN agrees to release protocol royalty-free
- **1994** Andreessen forms Netscape
- **1994** W3C formed, standardizing protocols, encouraging interoperability



# How does the web work?



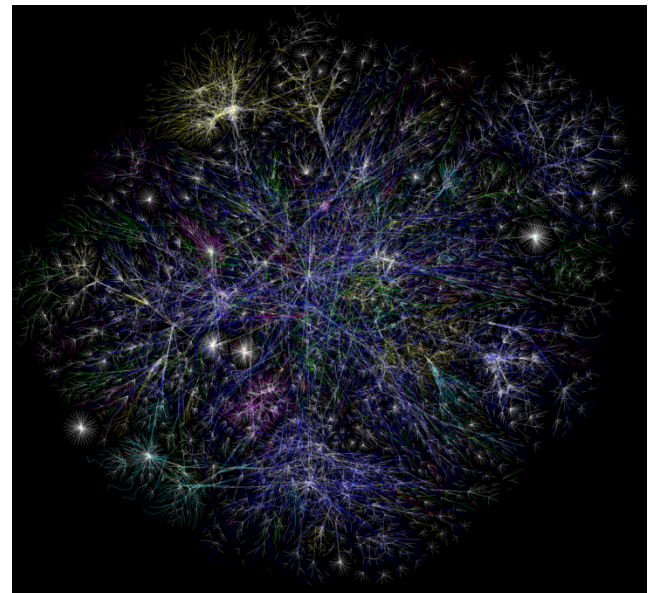
# What is the web?

- "The largest human information construct in history. The web is transforming society."  
– Web Science Trust

1989



2005



# How big is the web?

- "We've known it for a long time: the web is big."

– Google, <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>

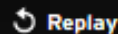
Year	Unique URLs
1998	26 million
2000	1 billion
2004	8 billion
2008	1 trillion

- Human population, ~7 billion
  - 143 pages / person
- 1 billion Google search queries a day



<http://funquisha.deviantart.com/art/ORIGAMI-ZEBRA-for-real-205044992>

## NEWS TECHNOLOGY

[Home](#) | [US & Canada](#) | [Latin America](#) | [UK](#) | [Africa](#) | [Asia](#) | [Europe](#) | [Mid-East](#) | [Business](#) | [Health](#) | [Sci/Environment](#)

Replay

THE FUTURE IS NOW.  
LET'S PUT IT TO WORK.



Business, Art  
Technology

26 January 2011 Last updated at 12:59 ET

659 Share

## Confusion over Egyptian blocks on web protest tools

**Confusion surrounds the use of web tools such as Twitter and Facebook that have been used by protesters in Egypt to co-ordinate action.**

The Egyptian government denied taking any action to restrict use of the web, saying it respected freedom of expression.

However, Twitter said it is being blocked but said many people have found ways round the restrictions.



Egyptian anti-government activists clashed with police in defiance of a ban on protests



lease for \$99/month  
for 36 months.

LEGAL



> find a

26 January 2011 Last updated at 05:39 ET

1.3K

Share



## Twitter and web video site face clampdown in Egypt

By Jonathan Fildes

Technology reporter, BBC News

**Egypt appears to have clamped down on web services, such as Twitter, that have been used to help organise anti-government protests in Cairo.**

Twitter confirmed that its service has been blocked in Egypt on Tuesday from around 1600GMT.

A Swedish mobile video site called Bambuser also reported that it had been blocked around the same time.



Riot police used water cannons to disperse demonstrators

## De-risking the cloud with end-to-end SLAS Smart Wired Cloud Seminar

# Egypt blocks Internet access amid protests

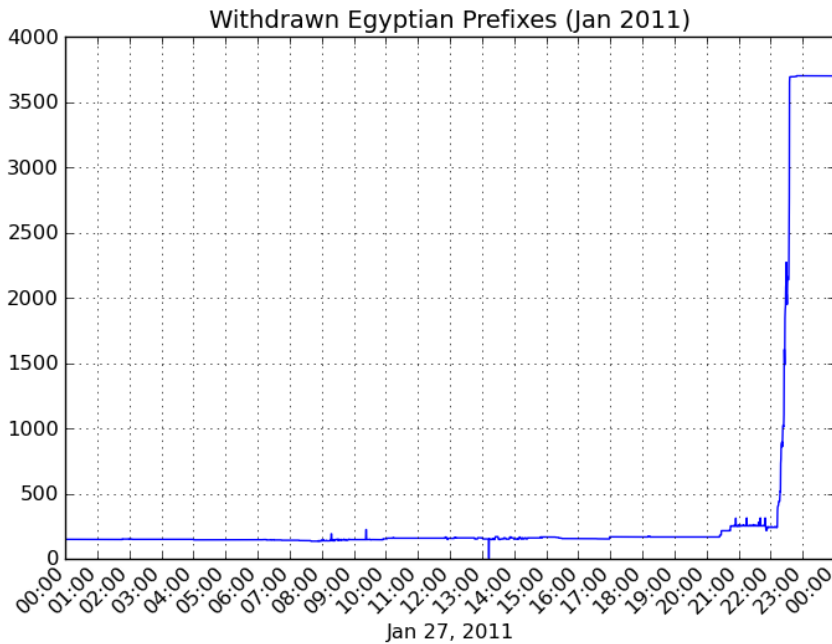
28 JANUARY 2011 Daniel Shane



**Government orders telcos to block web access as protestors take to the streets**

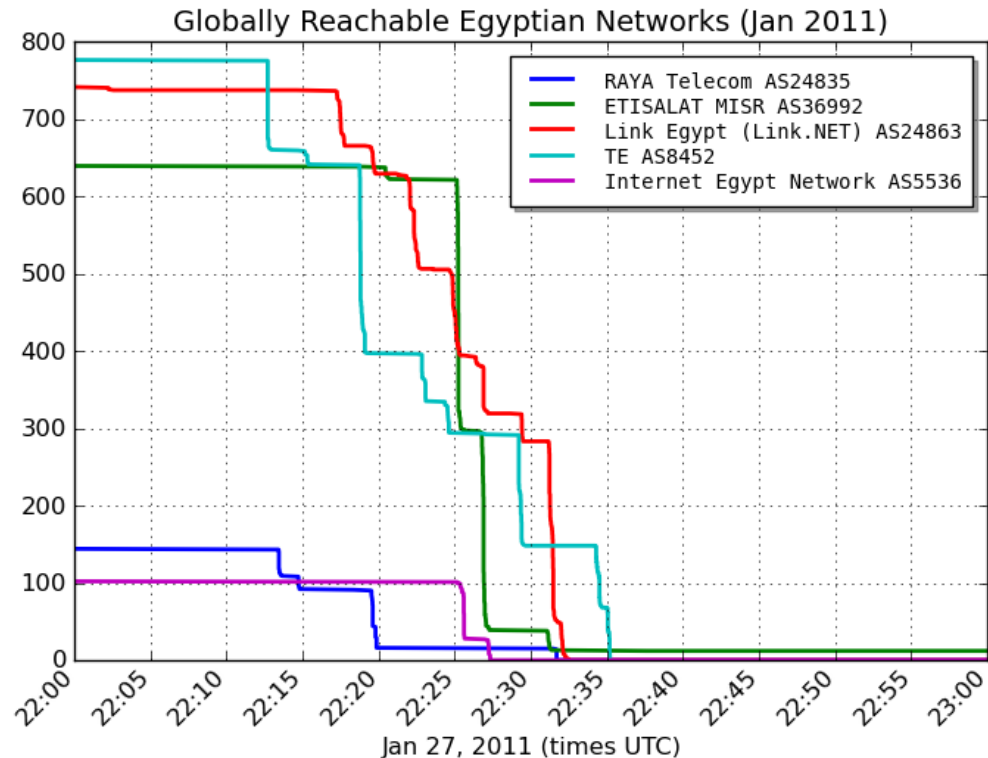
The Egyptian government has called on telecommunications providers in the country to block access to the Internet in response to widespread civil unrest.

Vodafone Egypt, one of the largest operators in the country not controlled by the state, today said it has disabled access following pressure from authorities.



"What happens when you disconnect a modern economy and 80,000,000 people from the Internet?"

<http://www.renesys.com/blog/2011/01/egypt-leaves-the-internet.shtml>





Demonstrators in Cairo's Tahrir Square, February 8, 2011

# What is "web science"?

- **Science 2006**
  - MIT, University of Southampton

## **Creating a Science of the Web**

**Tim Berners-Lee<sup>1</sup>, Wendy Hall<sup>2</sup>, James Hendler<sup>3</sup>, Nigel Shadbolt<sup>2</sup>, Daniel J. Weitzner<sup>1</sup>**

<sup>1</sup>Computer Science and Artificial Intelligence Laboratory, MIT, USA.

<sup>2</sup>School of Electronics and Computer Science, University of Southampton, UK.

<sup>3</sup>Computer Science Department, University of Maryland, USA.

Since its inception, the World Wide Web has changed the ways scientists communicate, collaborate, and educate. There is, however, a growing realization among many researchers that a clear research agenda aimed at understanding the current, evolving, and potential Web is needed. If we want to model the Web; if we want to understand the architectural principles that have provided for its growth; and if we want to be sure that it supports the basic social values of trustworthiness, privacy, and respect for social boundaries, then we must chart out a research agenda that targets the Web as a primary focus of attention.

# What is "web science"?

- "Web science...is inherently interdisciplinary and integrates computer and information sciences with a multitude of disciplines including sociology, economics, political science, law, management, language and communication, geography and psychology."

# This course

- **Focus on technical side**
  - Learn interesting technologies
  - Build stuff
- **Objectives** (from the course catalog):
  - Understand and the resources, **data formats**, and **protocols** used in the Internet
  - Build **robust and load balanced client server** software
  - Understand and use public/private key **cryptography**
  - Understand **state management** on the Internet

# Tentative topics

- Technologies

- LAMP: Linux Apache MySQL PHP
- Client-side scripting, e.g. JavaScript
- Server-side scripting, e.g. PHP
- HTML5
- Mobile device development (iOS or Android)

- State management

- Storing state at the client or server
- Cookies, session variables, query string, databases

- Dynamic content generation

- CGI, server-side scripting, web server modules, servlets



# Tentative topics

- Web services, cloud computing
  - SOAP, REST
  - JSON
  - Ajax
- Crowdsourcing
  - Amazon Mechanical Turk
- Availability
  - Providing fault tolerance
  - 99.999%
- Scalability
  - Scaling to million/billions of hits

# Tentative topics

- Performance
  - How to measure, e.g. with tools like Apache Benchmark
  - How to improving, e.g. using tools like memcached
- Data formats
  - Audio, images, video
- Compression
  - The only thing making Netflix/YouTube work on today's networks
- Encryption
  - Symmetric encryption, e.g. DES, 3DES, AES
  - Public/private encryption, e.g. SSL
  - Authentication

# Paper details

- Paper:
  - Due 10PM, Friday 4/25
  - 1500 words minimum, 3000 words maximum
  - Technical paper with references:
    - Abstract
    - Introduction
    - Content section(s)
    - Conclusions
    - References
  - Teach me something about the subject
  - No substantial overlap with covered topics
  - 10% of your grade
  - No late days

# Presentation details

- Presentations:
  - Last week of class (and a bit)
  - ~15 minutes + questions
  - Audio/visual aids and demos as appropriate
  - 5% of your grade
  - No late days

# Past paper/presentation topics

- Massive data centers
- Defending against hackers
- Bitcoins
- Geolocation
- ...