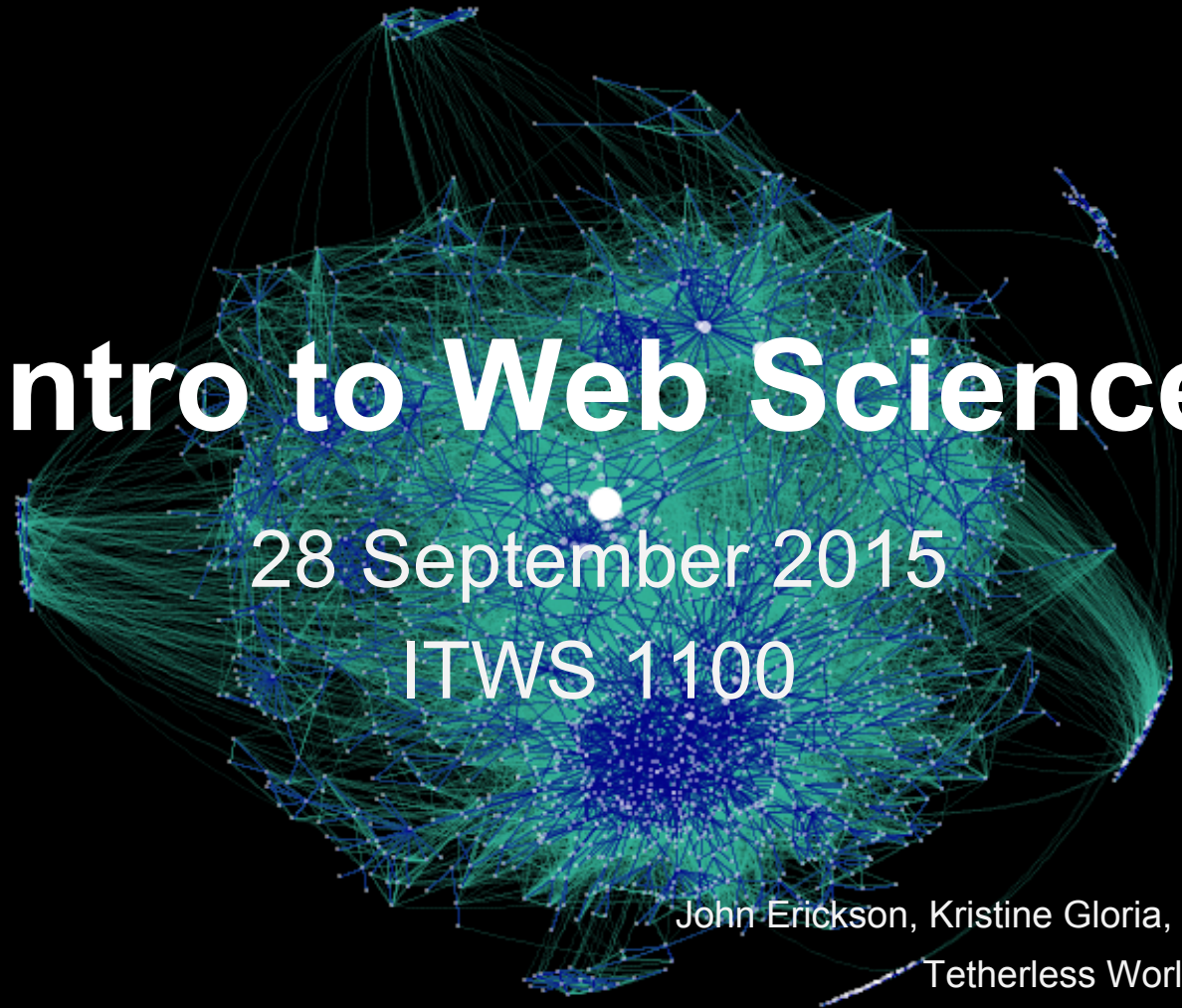


Intro to Web Science

28 September 2015

ITWS 1100

John Erickson, Kristine Gloria, Lindsay Poirier

Tetherless World Constellation

Rensselaer Polytechnic Institute

Agenda

1. A Science of The Web *and why it matters*
2. Web Architecture/Engineering the Web
3. Measuring the Web
4. The Web Science Method
5. Social Aspects of the Web
    a. Evolution of methodology
    b. Hurdles of incorporating the “social”
    c. Why humans aren’t just "nodes" in a network
6. Web and other Governance

What is Web Science?

- Positions the World Wide Web as an object of scientific study unto itself
- Recognizes the Web as a transformational, disruptive technology
- Its practitioners focus on understanding the Web...
  - ...its components, facets and characteristics

The Web Science Method:

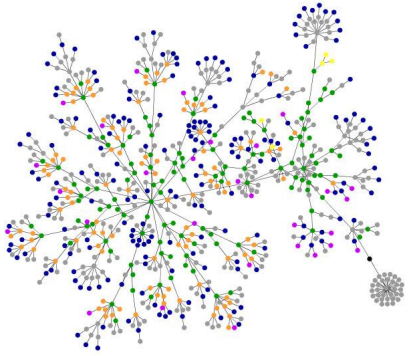
The process of designing things in a very large space...

What does Web Science ask?

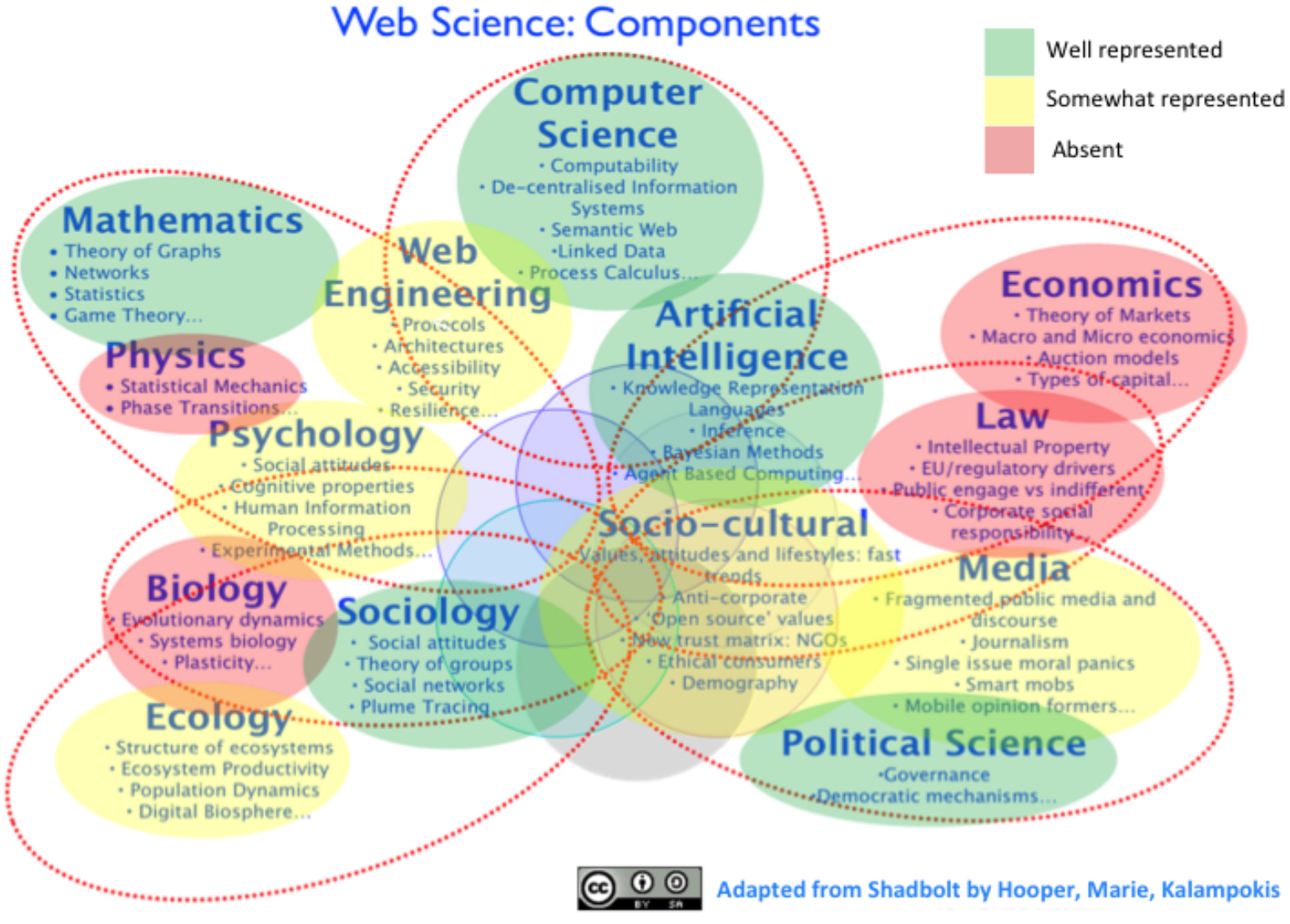
- What processes have driven the Web's growth, and will they persist?
- How does large-scale structure emerge from a simple set of protocols?
- How does the Web function as a socio-technical system?
- What drives the viral uptake of certain Web phenomena?

What might fragment the Web?

What is the Web?



- "The Web is not a thing..."
- **Continuously changing** due to coordinated *and* conflicting processes
- An **evolving large-scale structure** dependant on static *and* emerging protocols
- A **socio-technical system** that reflects *and* obfuscates social and technical structures
- The Web *always* goes where we *allow* it to go...
  - ...but seldom where we *want* or *expect* it to go!



Clare Hooper, et.al. http://bit.ly/R813sC

Web Architecture

It's really quite simple! ;)

- A standard system for **identifying** resources
- Standard formats for **representing** resources
- A standard protocol for **exchanging** resources

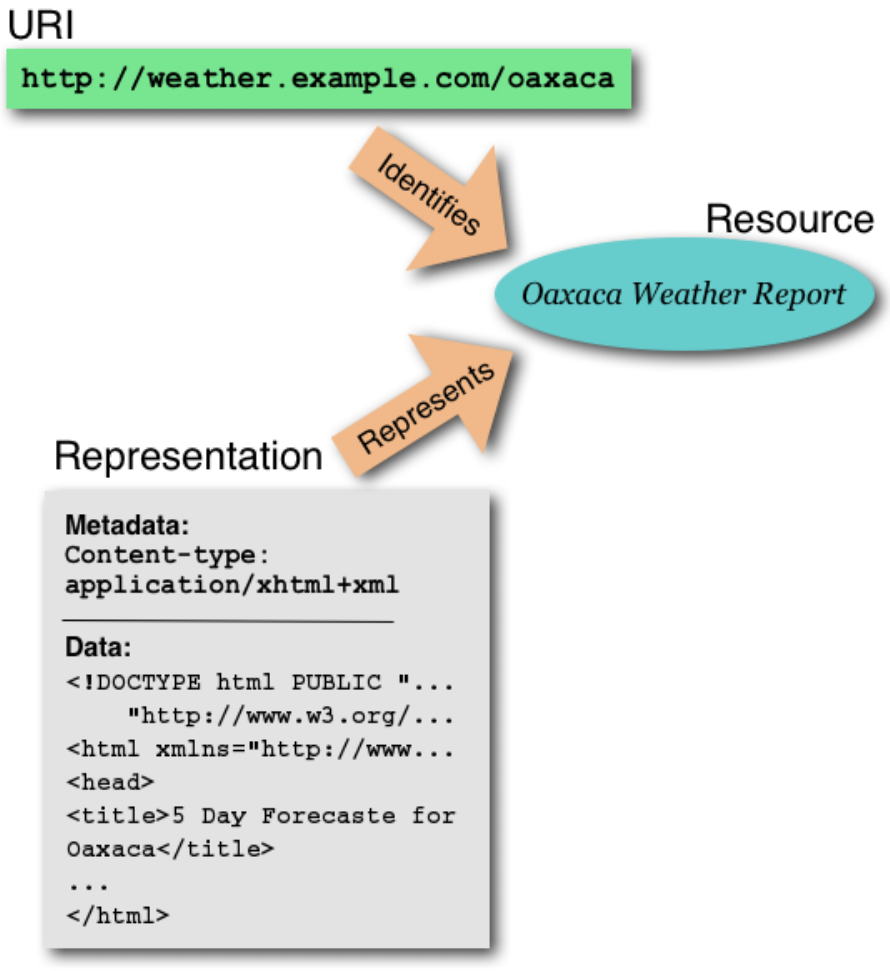
Web Architecture

It's really quite simple! ;)

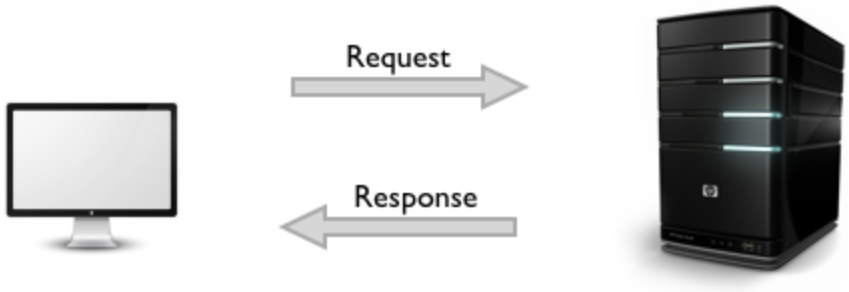
- A standard system for **identifying** resources
- Standard formats for **representing** resources
- A standard protocol for **exchanging** resources

Relevant core standards:

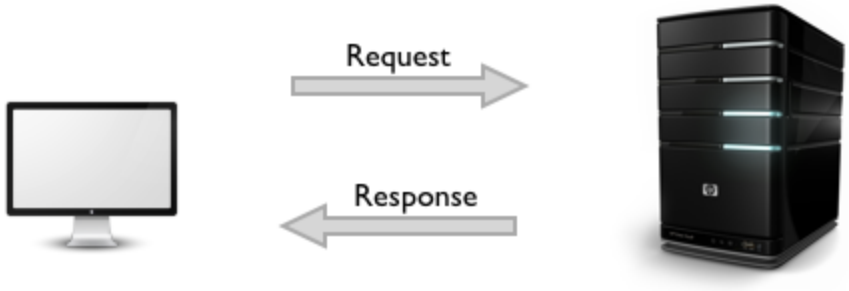
- **URI** (URL): Universal Resource Identifiers
- **HTML**: Hypertext Markup Language
- **HTTP**: Hypertext Transfer Protocol

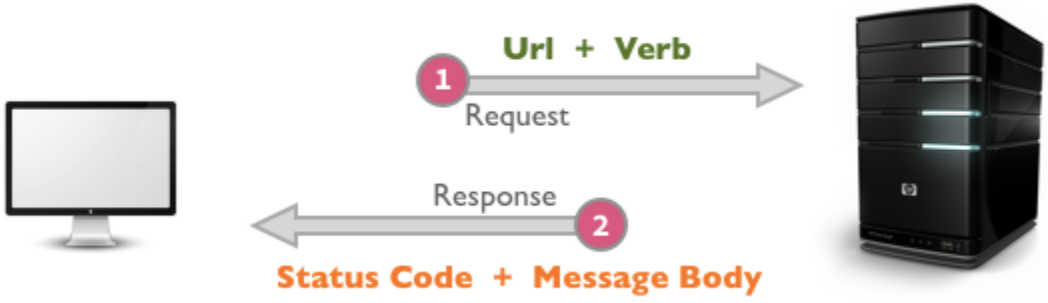


Architecture of the World Wide Web, Volume One http://www.w3.org/TR/webarch/

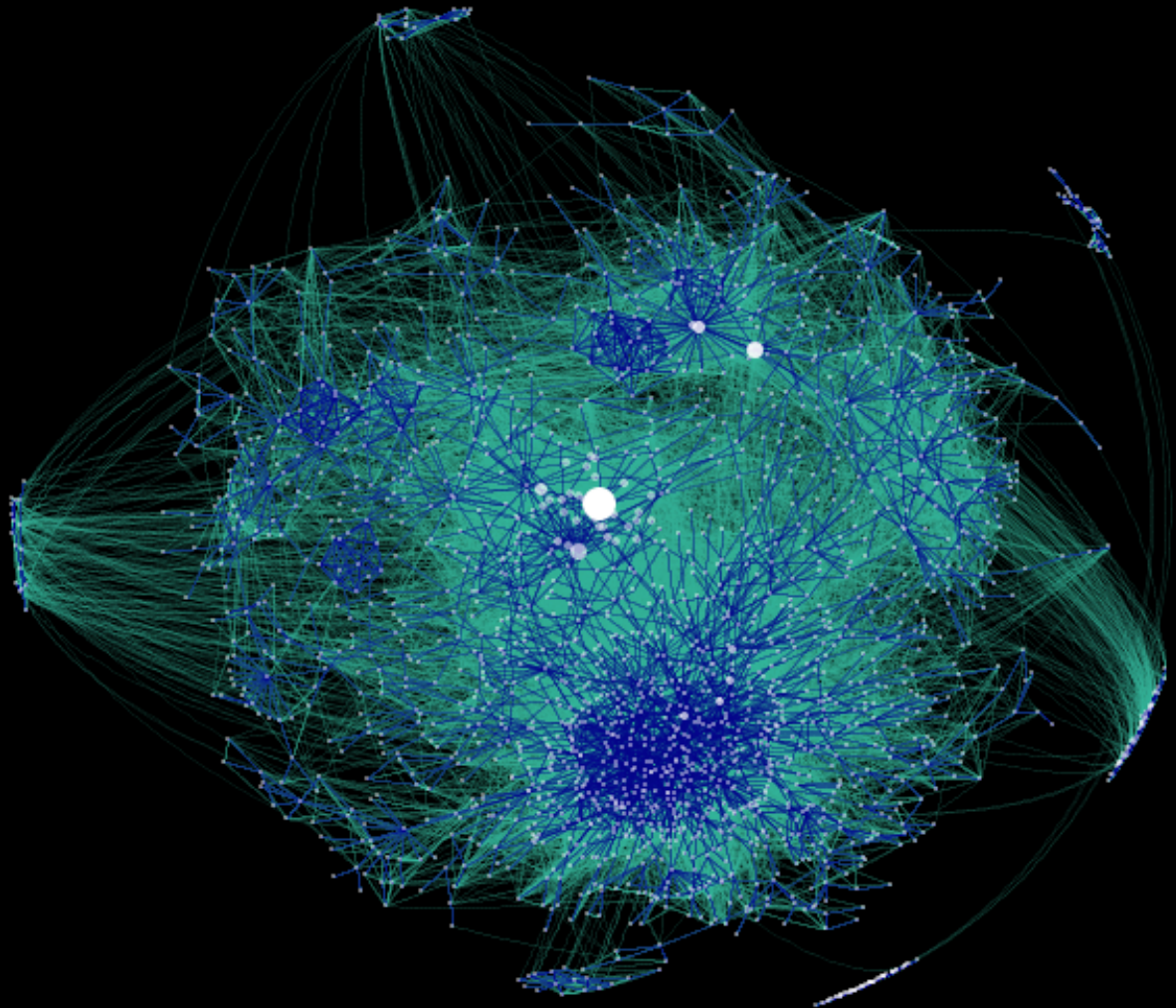


Figures from "HTTP Basics" http://bit.ly/1qX9Y2h

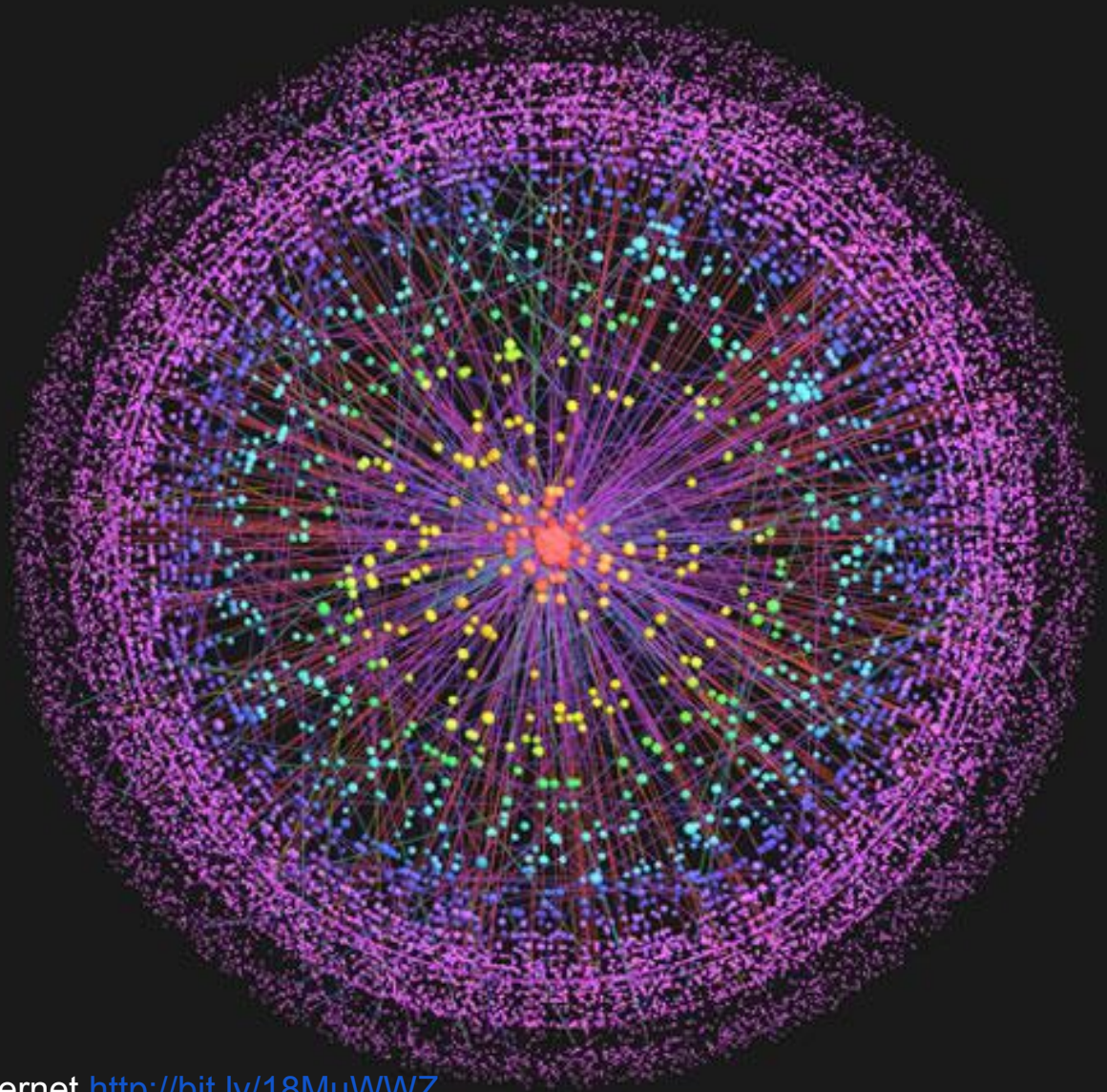




Figures from "HTTP Basics" http://bit.ly/1qX9Y2h



Data Mining: Mapping the Blogosphere http://bit.ly/18MuXdD



Mapping the Internet http://bit.ly/18MuWWZ

Identifying Resources (1)

- A global identification system is essential
  - to share information about resources
  - to reason about resources
  - to modify or exchange resources
- **Resources:** anything that can be **linked to**, spoken of
  - Documents, cat videos, people, ideas...
- Not all resources are "on" the Web
  - They might be *referenced from* the Web...
  - ...while not being *retrievable from it*
  - These are (so-called) "information resources"

Identifying Resources (2)

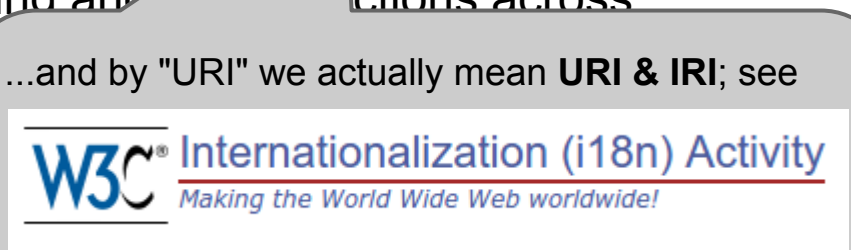
- A global standard is required; the **URI*** is it
- Others systems are possible...
  - ...but added value of a **single global system** of identifiers is high
  - Enables linking, bookmarking and other functions across heterogeneous applications
- How are URIs used?
  - **All resources** have URIs associated with them
  - URIs identify **single resources** in a context- independent manner
  - URIs act as names and (usually) addresses
  - In general URIs are "opaque"

Uniform Resource Identifier (URI): Generic Syntax (RFC 3986) http://www.ietf.org/rfc/rfc3986.txt

Identifying Resources (2)

- A global standard is required; the **URI*** is it
- Others systems are possible...
  - ...but added value of a **single global** **m** of identifiers is high
  - Enables linking, bookmarking and ctions across heterogeneous applicatio
- How are URIs used?
  - **All resources** have URIs
  - URIs identify **single reso**
  - URIs act as names and (
  - In general URIs are "opa

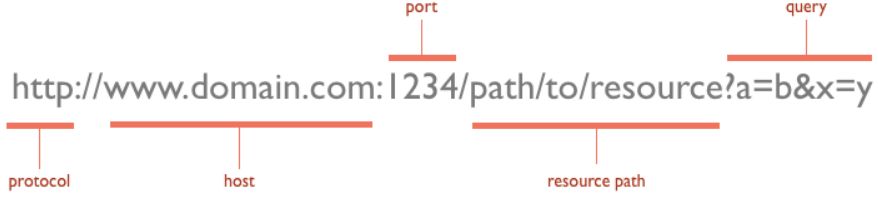
...and by "URI" we actually mean **URI & IRI**; see



Uniform Resource Identifier (URI): Generic Syntax (RFC 3986) http://www.ietf.org/rfc/rfc3986.txt

Identifying Resources (4)

- "URIs *identify* and URLs *locate...*"
  - ...and also identify
- URLs are URIs aligned with protocols
  - URLs include the "access mechanism" or "network location", e.g. http:// or ftp://
  - How to "dereference" the URI and retrieve the thing
- URL examples
  - **ftp://**ftp.is.co.za/rfc/rfc1808.txt
  - **http://**www.ietf.org/rfc/rfc2396.txt
  - **mailto:**John.Doe@example.com
  - **telnet://**192.0.2.16:80/

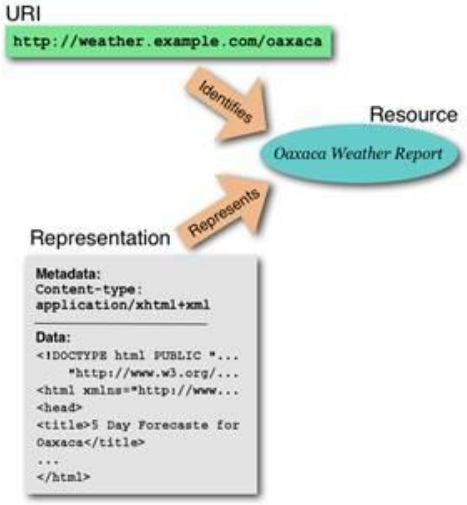


Uniform Resource Identifier (URI): Generic Syntax (RFC 3986) http://www.ietf.org/rfc/rfc3986.txt

Representing Resources (1)

- Resources are (usually) manifest as digital files
- The Web recognizes a (growing) set of file formats
  - The original and workhorse is HTML...
  - ...but there are many others
- Retrievable resources on the Web serve multiple purposes
  - Resources **encode** information and data
  - Resources **aggregate links** to other resources
- This is what makes **The Web(tm)** a **web...**

Resources (nodes) aggregate links to other resources to create a web



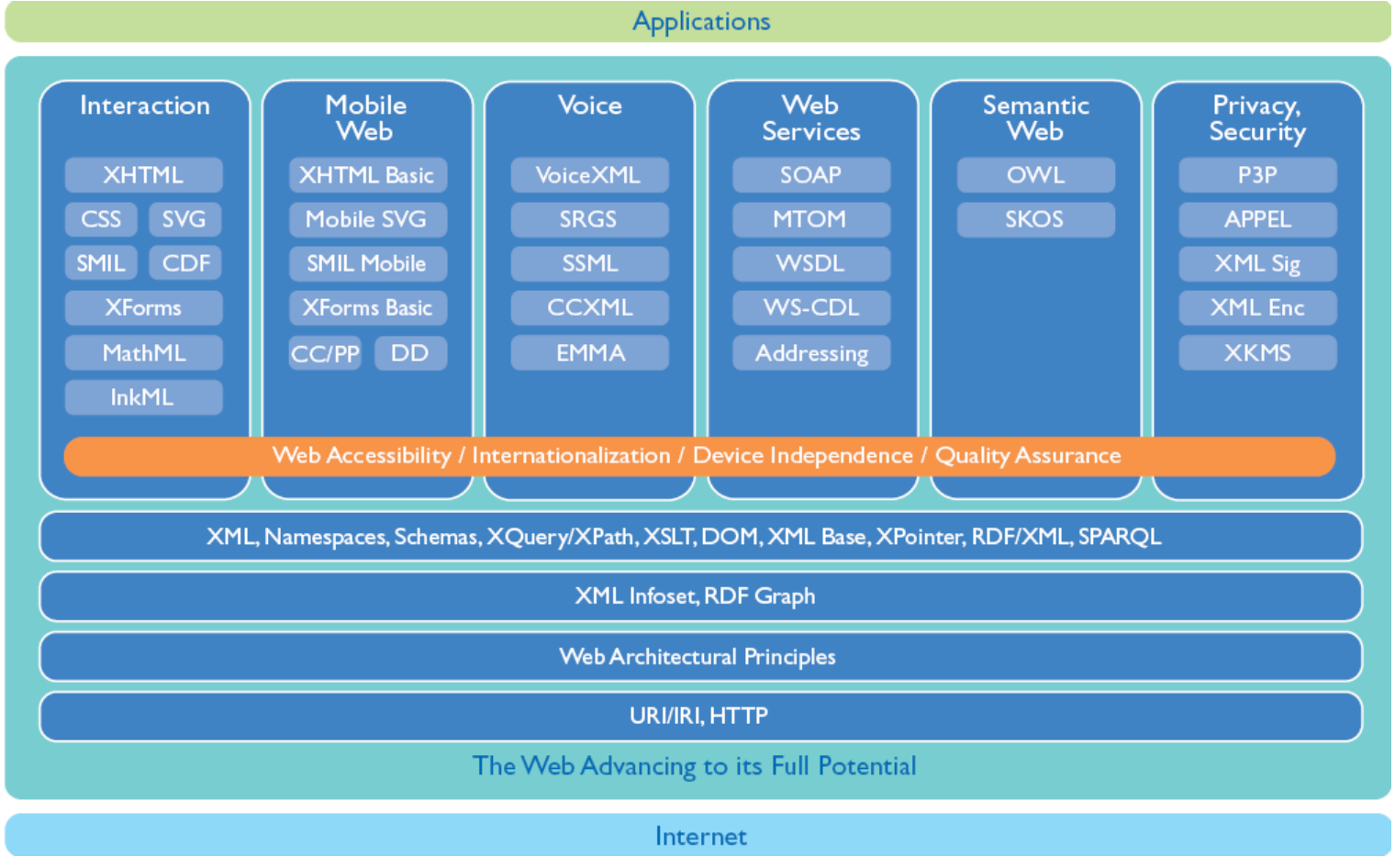
Retrieving Resources (1)

Review:

- URIs that reference *retrievable resources* -- URLs -- must specify a *protocol* for retrieval
- The original and most common Web protocol is **HTTP**
- Specialized protocols are possible...

  …but those resources may appear "off the grid..."

URIs, HTTP, many formats...



Principles for creating a *healthy* Web

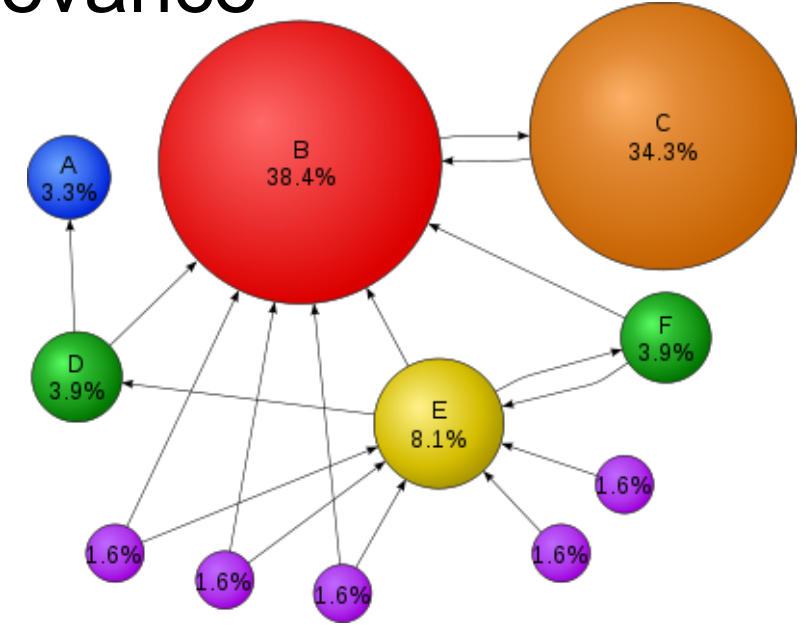
- Use URIs as names for **things**
- Use HTTP URIs so people can **look up** those names
- When someone looks up a URI, **return useful information**
  - ...and use the **standards** to do it
- Include **links** to other URIs, so the "consumer" can discover more things
  - "Consumer" as in "eater of pizza" {people | applications}

Why is linking important???

Tim Berners-Lee http://www.w3.org/DesignIssues/LinkedData.html

Implications of a well-connected Web: Google PageRank

- Links to other nodes as a "vote" of quality and/or relevance

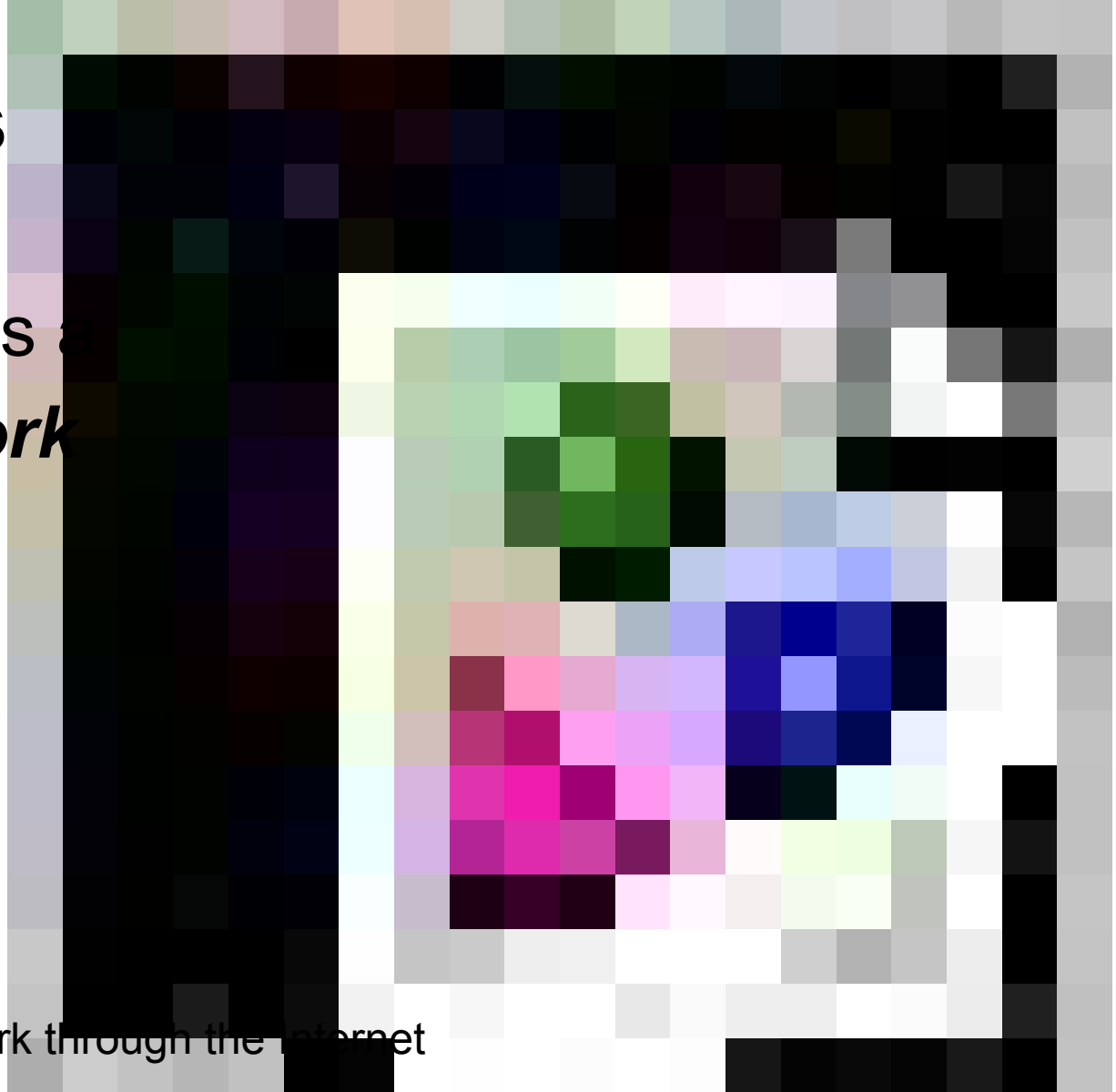


PageRank https://en.wikipedia.org/wiki/PageRank

Meas

Web as a

Network



Router network through the Internet

Measuring the Web

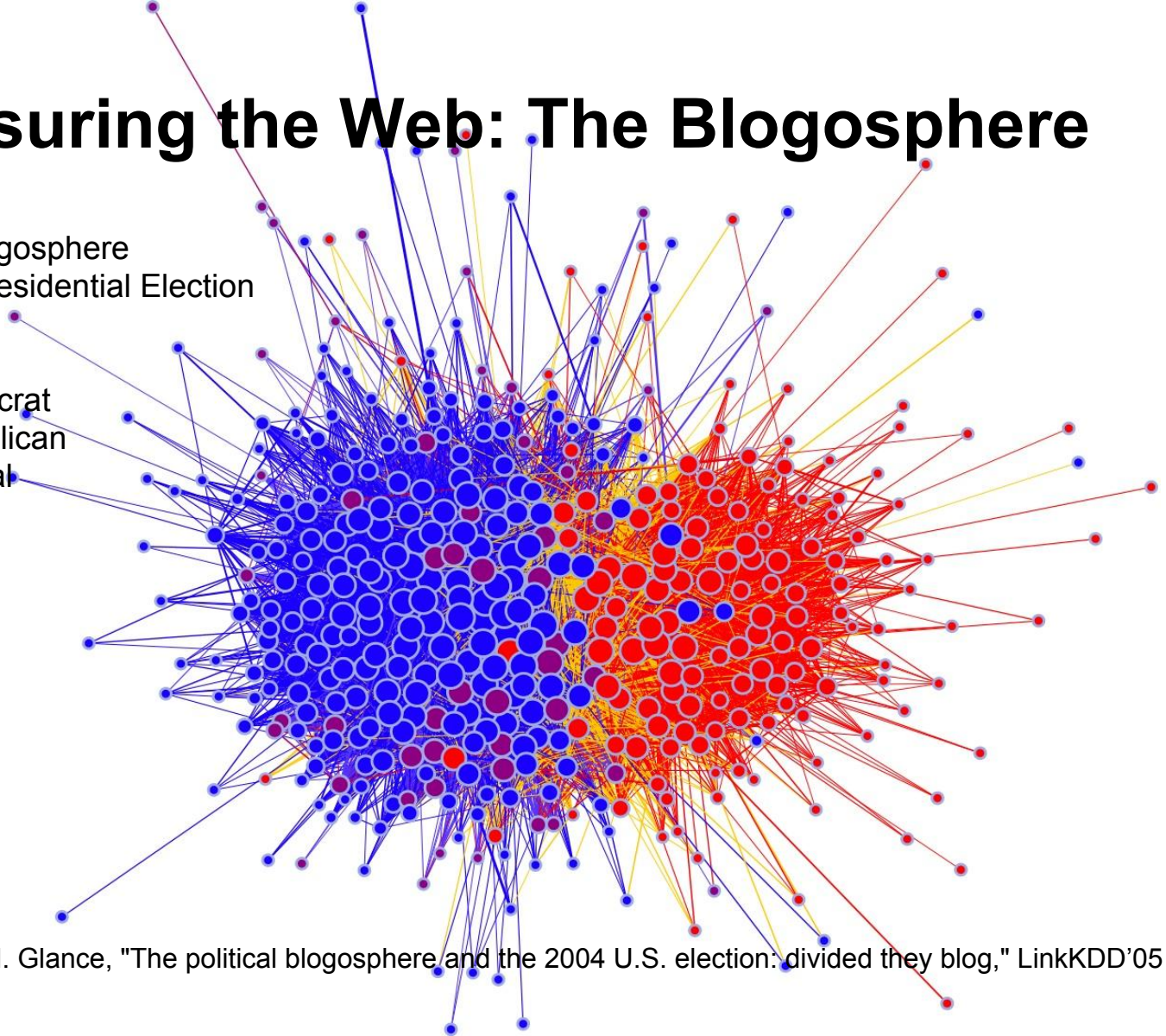
- The rich variety of networks "on" the Web
  - Router network
  - Web page network (linking via hyperlinks)
  - Document network (citation network on DBLP*, etc.)
  - Social networks: (linking via human relationships)
    - **Facebook:** friendship, comment-reply, tag, and all kinds of social relationship on Facebook
    - **Twitter:** follower, retweet, mention, reply, etc.
    - **Blogosphere**: friendship, visiting, comment, etc.
    - **LinkedIn:** colleague, classmate, etc.
    - **Crowdsourcing:** collaboration, co-worker, etc.
    - Other social media...

*The DBLP Computer Science Bibliography http://www.informatik.uni-trier.de/~ley/db/

Measuring the Web: The Blogosphere

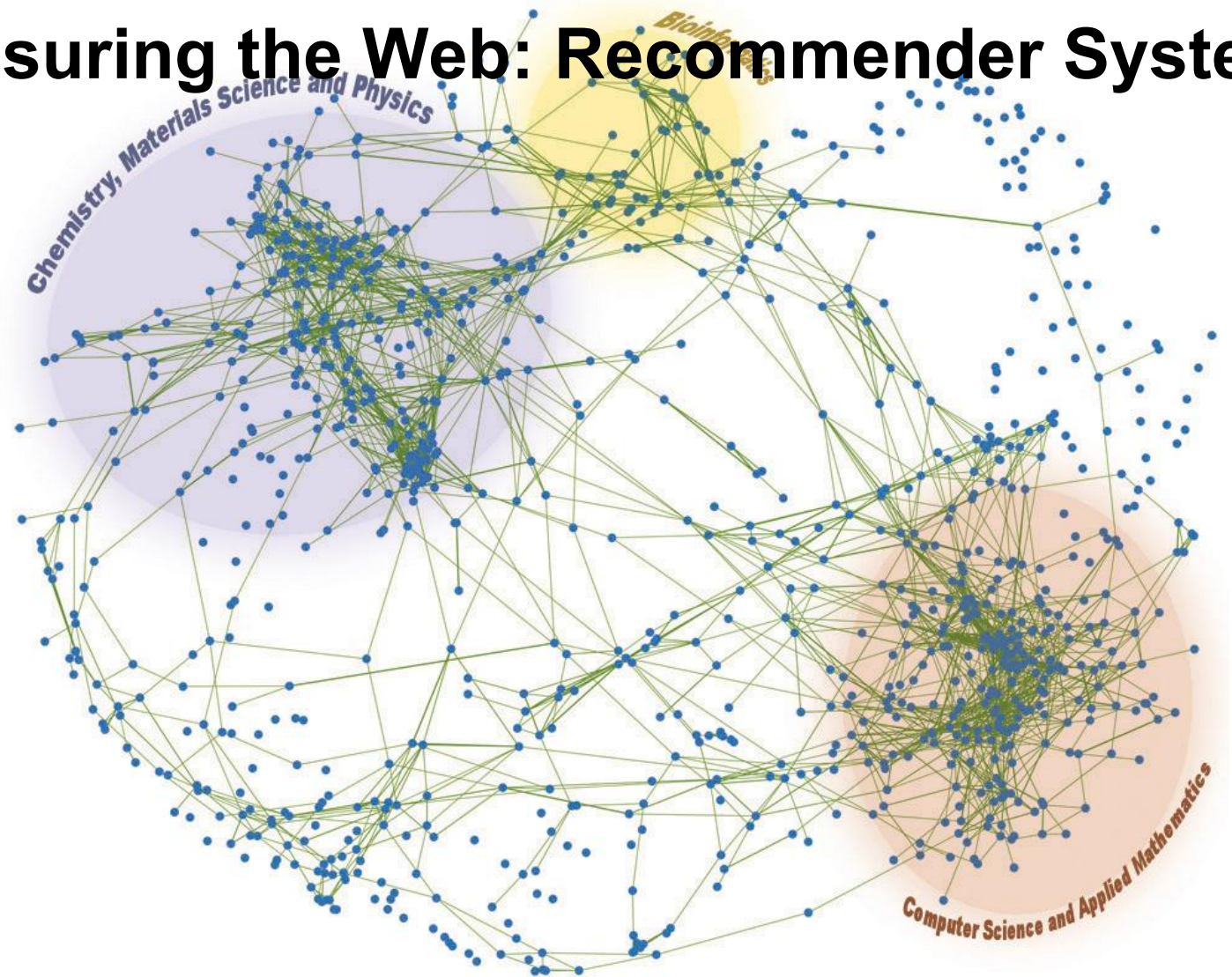
Political Blogosphere
2004 US Presidential Election

Bloggers:
Blue: Democrat
Red: Republican
Pink: Neutral

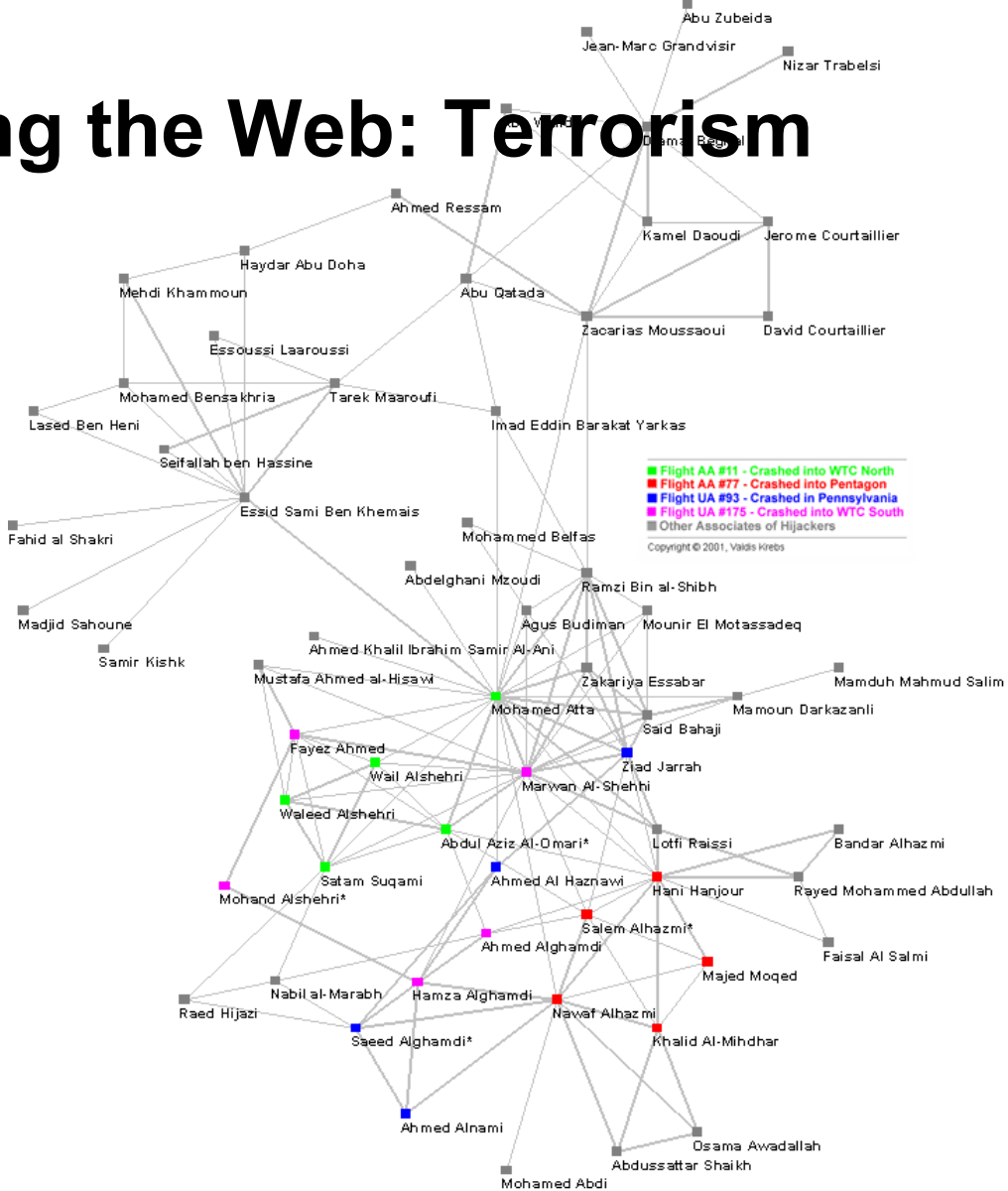


L. Adamic, N. Glance, "The political blogosphere and the 2004 U.S. election: divided they blog," LinkKDD'05

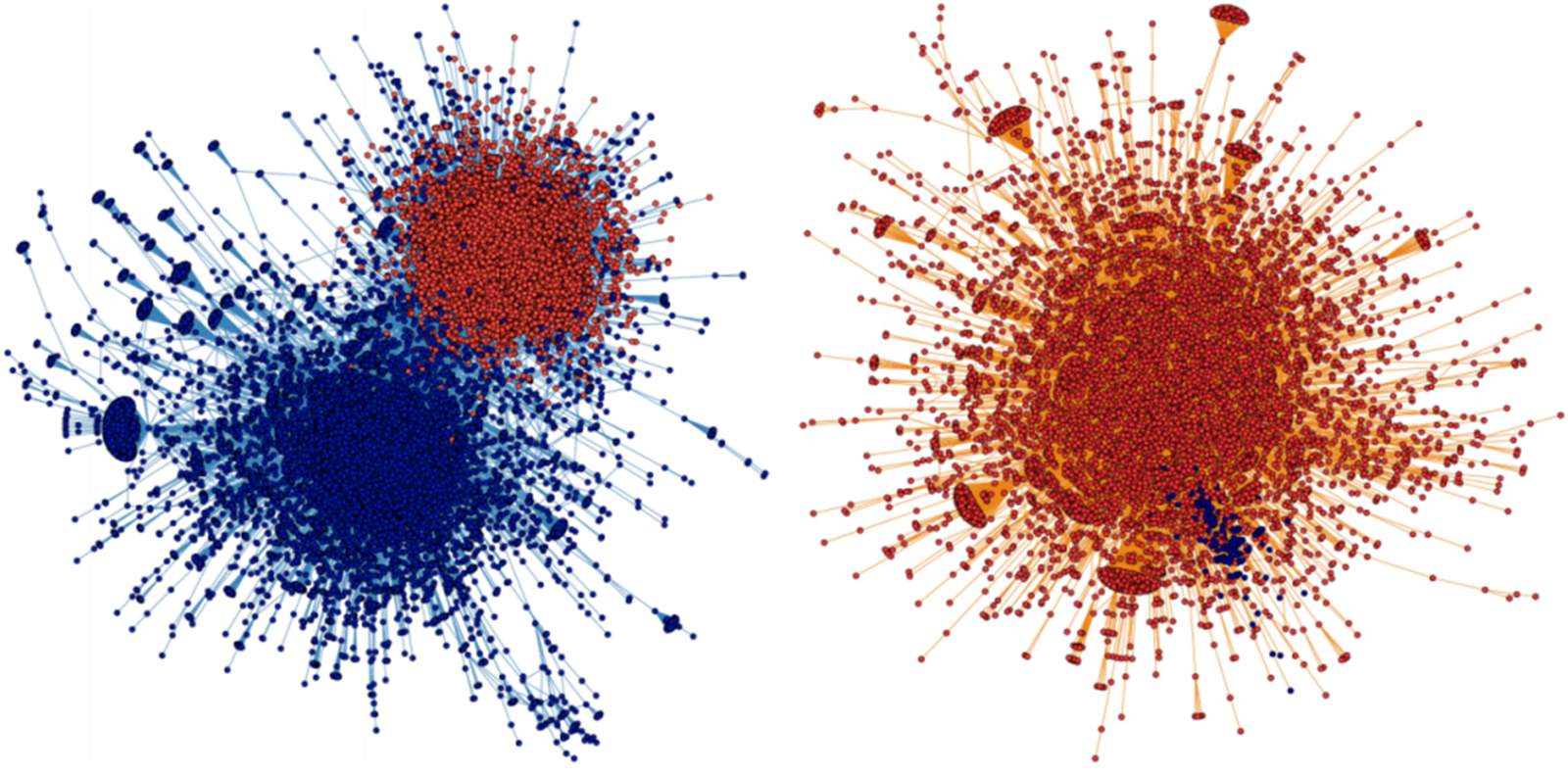
Measuring the Web: Recommender System



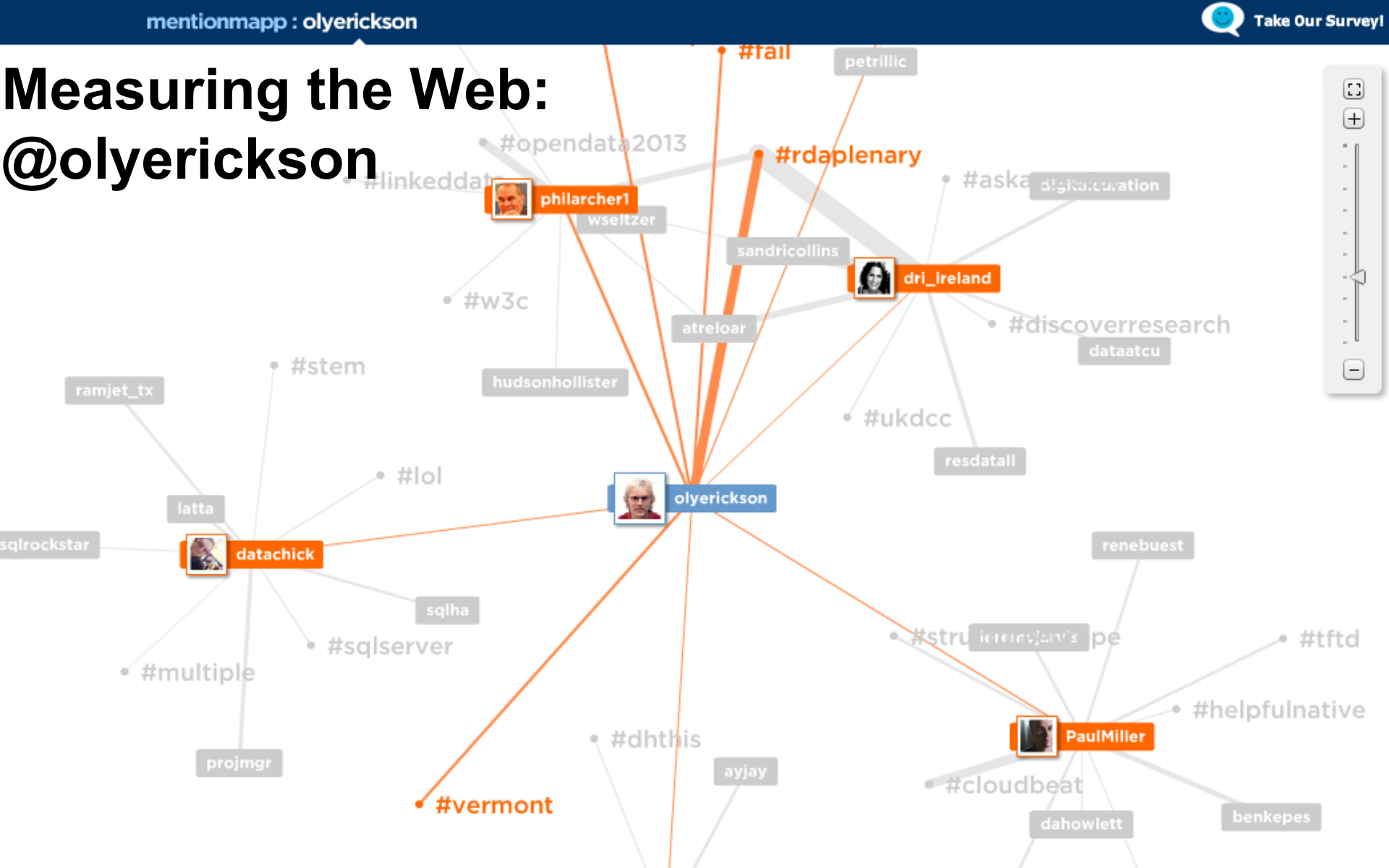
Measuring the Web: Terrorism



Measuring the Web: Twitter



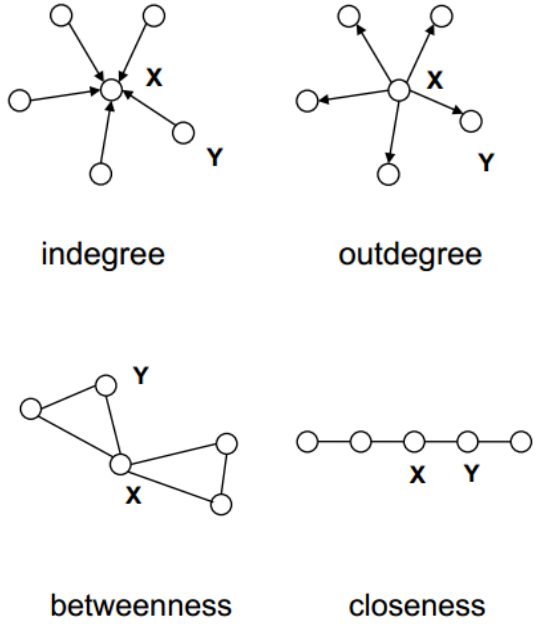
M. D. Conover, et al. Political Polarization on Twitter, ICWSM'11


mentionmapp : olyerickson
Take Our Survey!
Measuring the Web:
@olyerickson
#fail
petrillic
#opendata2013
#rdaplenary
philarcher1
wseltzer
sandricollins
dri_ireland
#w3c
atreloar
#discoverresearch
dataatcu
#stem
hudsonhollister
ramjet_tx
#ukdcc
resdatall
#lol
olyerickson
latta
datachick
sqlrockstar
renebuest
sqlha
#sqlserver
#multiple
#tftd
#helpfulnative
PaulMiller
#dhthis
projmgr
ayjay
#cloudbeat
#vermont
dahowlett
benkepes

Analyzing networks on the Web

We can measure...

- # of nodes
- # of edges
- "Diameter" and "radius"
- Network density
- Degree distribution
- Clustering coefficient
- Average shortest path length
- Strongly/weakly connected components
- Betweenness/Closeness centrality
- **Bow-tie structure**
- Community discovery
- Key nodes discovery
- etc...

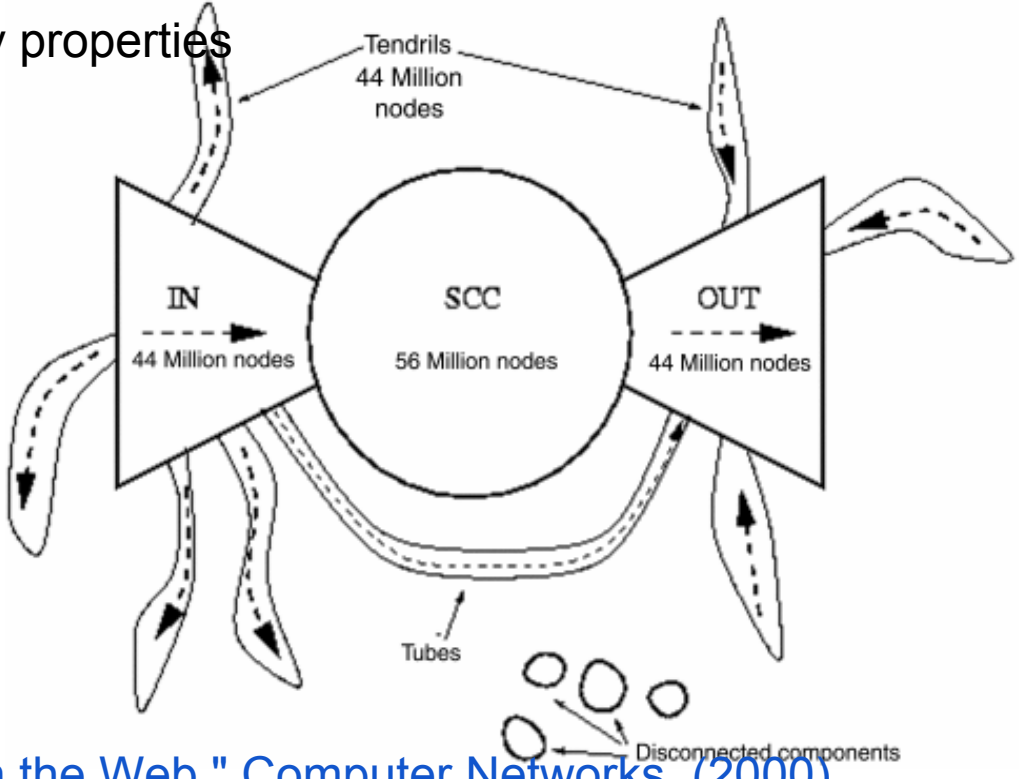


See e.g. Network Centrality

Measuring the Web: Bow Tie Structure

Overview of the Web's reachability properties

- Strongly Connected Core
- IN
- OUT
- Tendrils
- Tubes
- Disconnected



A. Broder et al. "Graph structure in the Web," Computer Networks, (2000)

Measuring the Web

- It's a “Small World” after all...
  - Most pairs of pages separated by small number of links
  - Almost always by fewer than 20 links
  - "Diameter" of central core is 28
    - very small compared to the size of the Web
  - Analysis suggests diameter will grow logarithmically with the size of the Web (ie slowly)
  - Diameter of social networks decreases over time
- Conclusion: The Web is “smaller” than we thought!
- “Six degrees of separation” verified in Social Web(s)

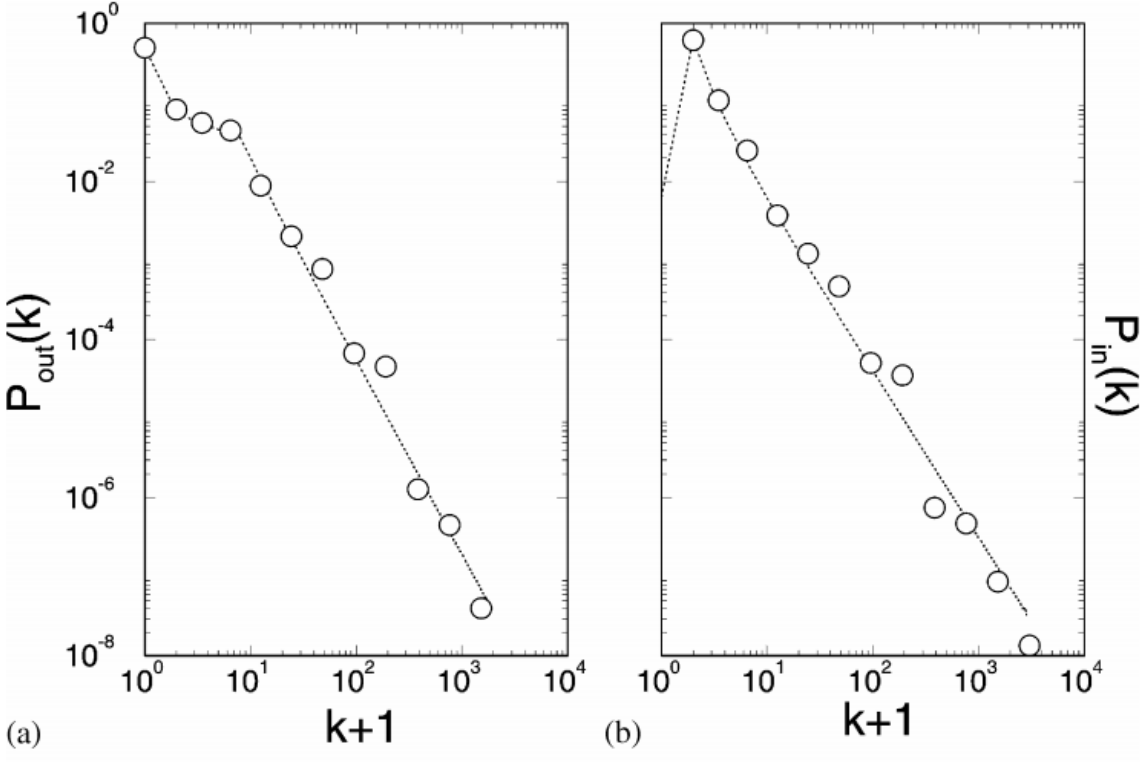
R. Albert, H. Jeong and A.-L. Barabasi, "Diameter of the World Wide Web," Nature 401 (1999). http://bit.ly/18atsYA

J. Leskovec, etc. "Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations," KDD (2005). http://bit.ly/1BPAYWh

Measuring the Web: Scale Free Properties

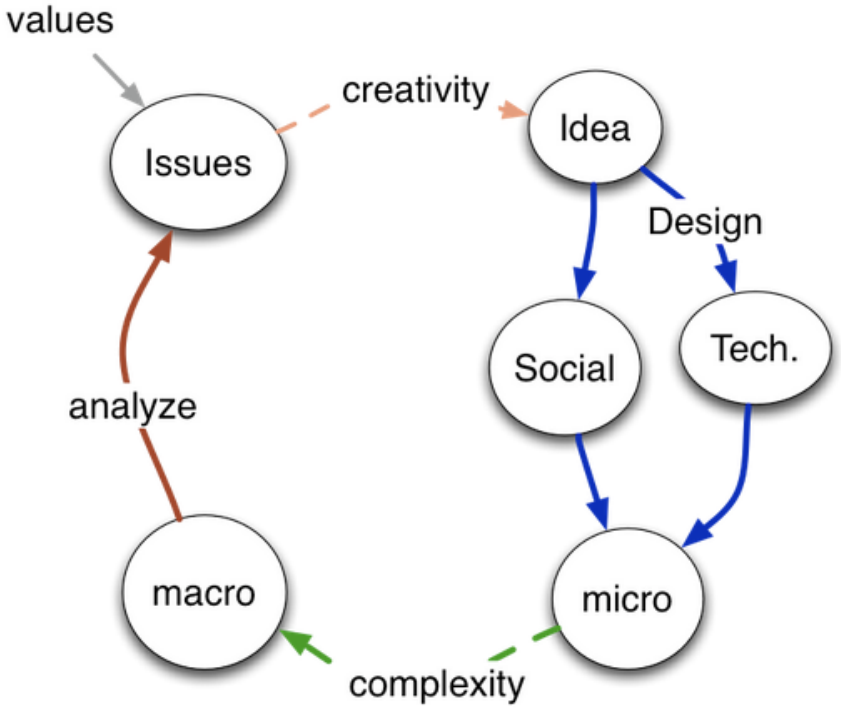
- Nodes of scale-free network aren't randomly or evenly connected
- Scale-free networks include many highly connected nodes that shape the way the network operates
- Ratio of highly-connected nodes to the number of nodes in the rest of the network remains constant as the network changes in size

$$P(k) \sim k^{-\gamma} \qquad 2 < \gamma < 3$$



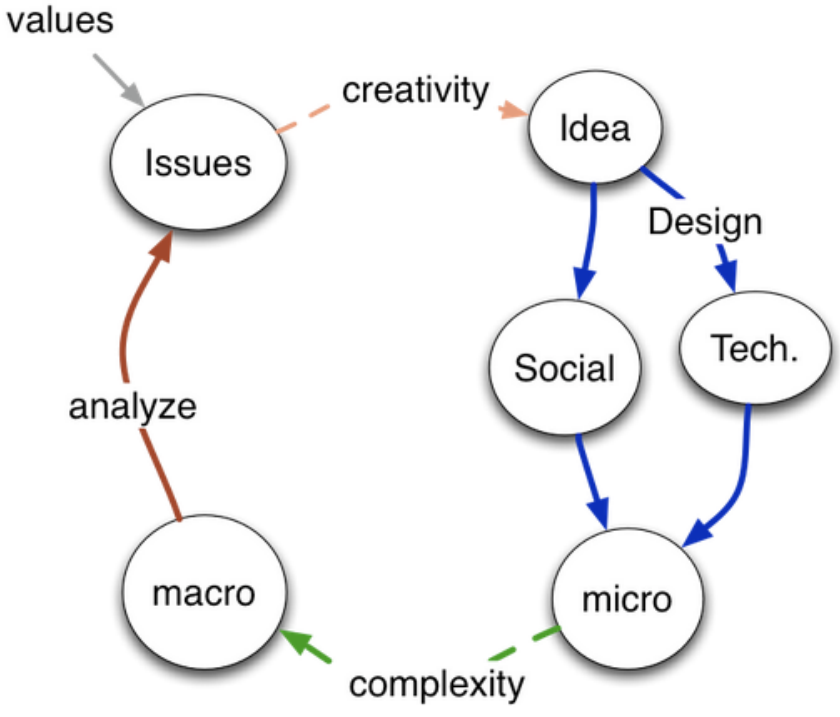
A live model: http://ccl.northwestern.edu/netlogo/models/PreferentialAttachment

The Web Science Method



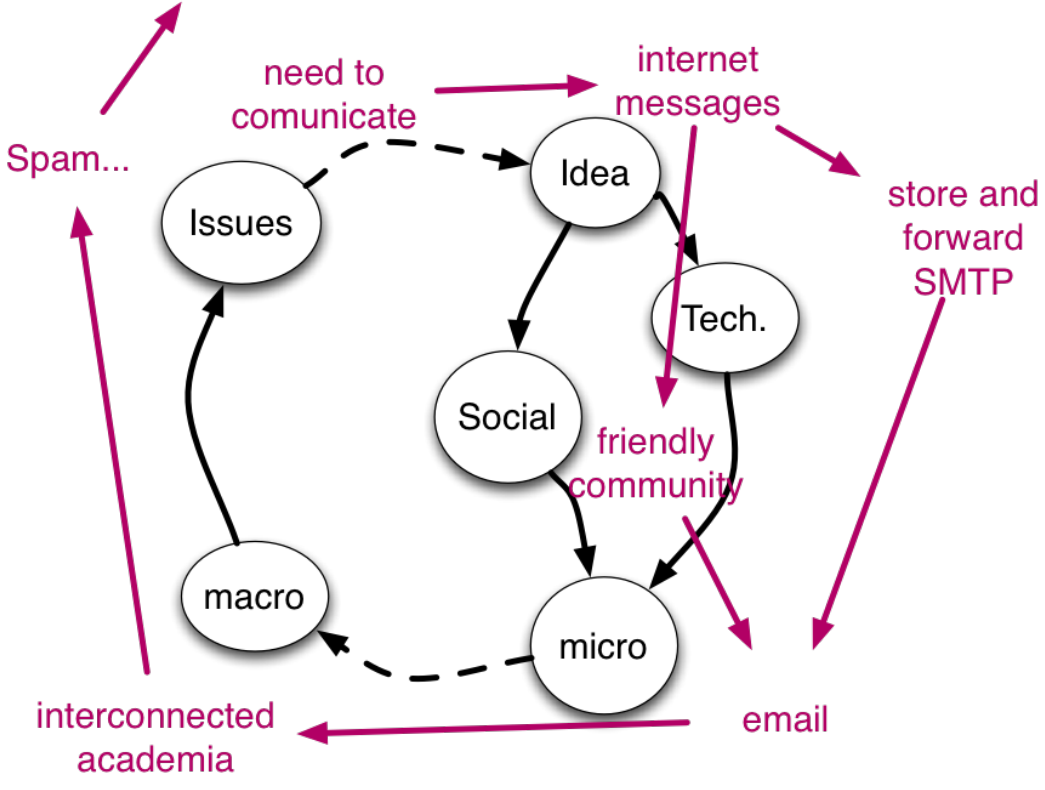
Berners-Lee, T. (2007). W3C. http://www.w3.org/2007/Talks/0509-www-keynote-tbl/#(10)

The Web Science Method



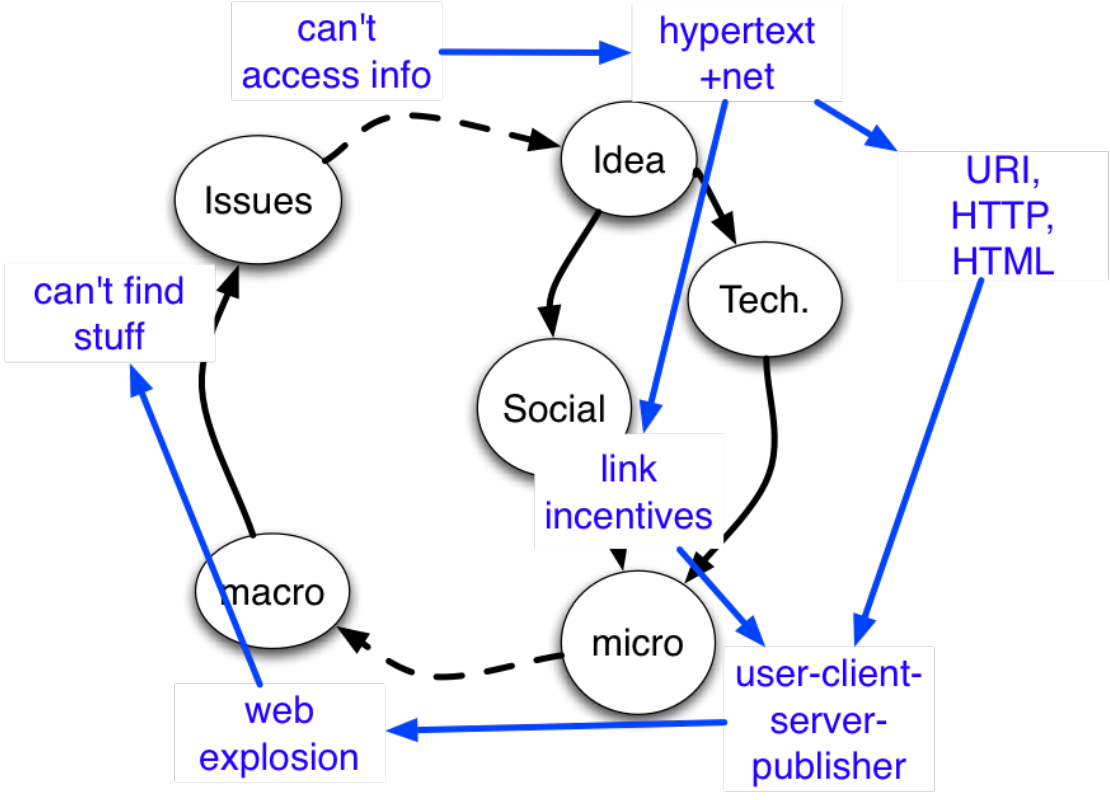
Berners-Lee, T. (2007). W3C. http://www.w3.org/2007/Talks/0509-www-keynote-tbl/#(10)

Applied to email...



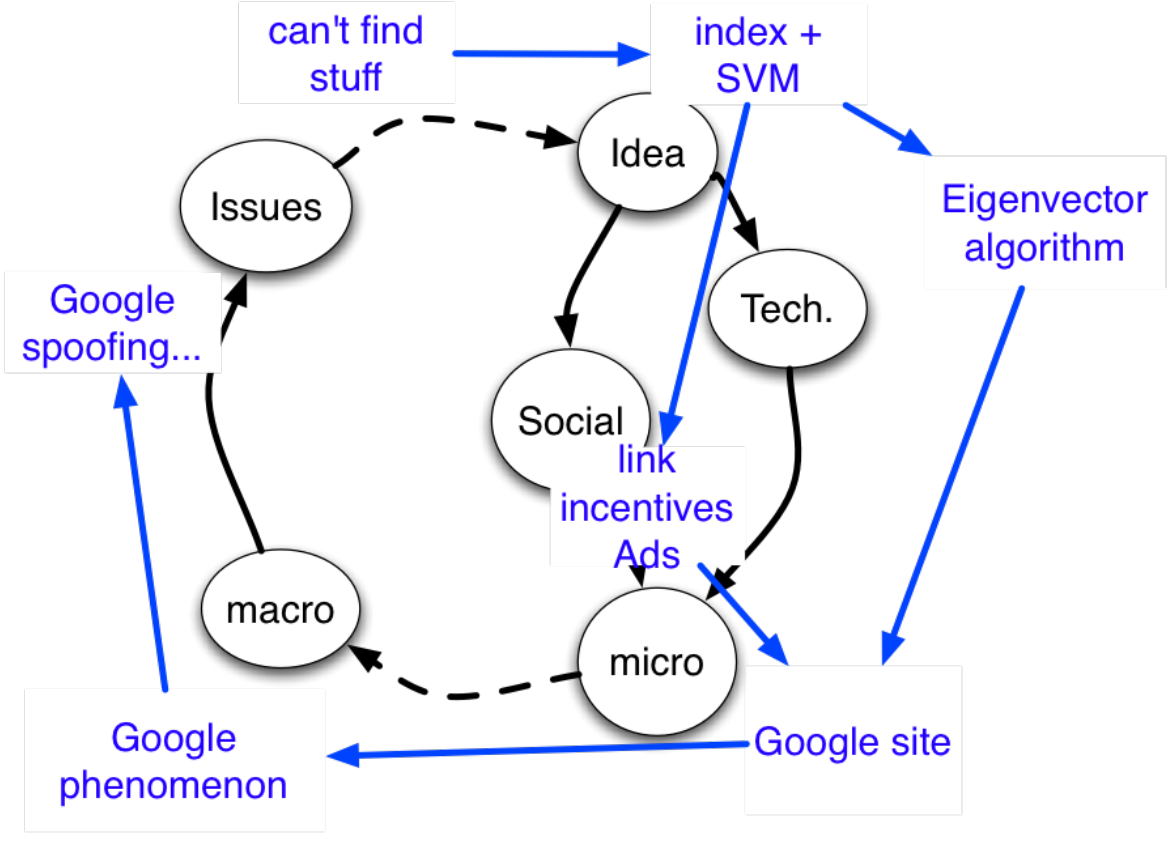
Berners-Lee, T. (2007). W3C. http://www.w3.org/2007/Talks/0509-www-keynote-tbl/#(16)

Applied to the original Web...



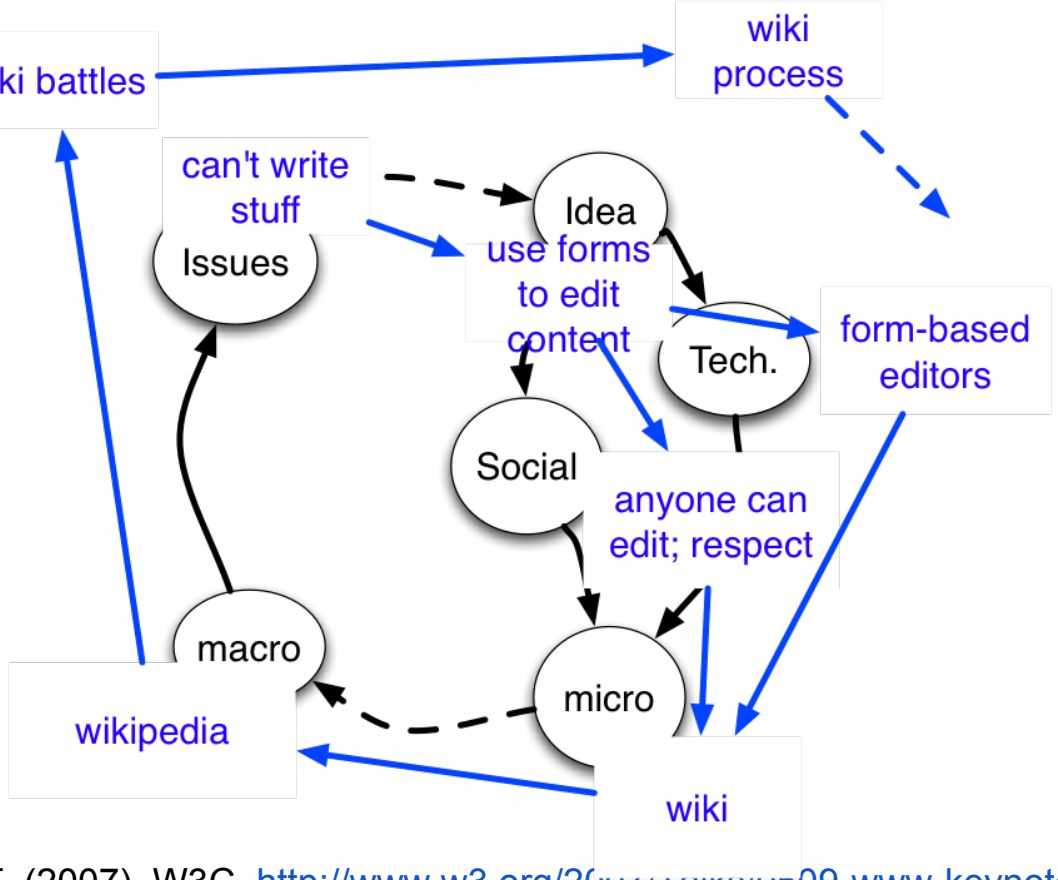
Berners-Lee, T. (2007). W3C. http://www.w3.org/2007/Talks/0509-www-keynote-tbl/#(18)

Applied to Google's Web...



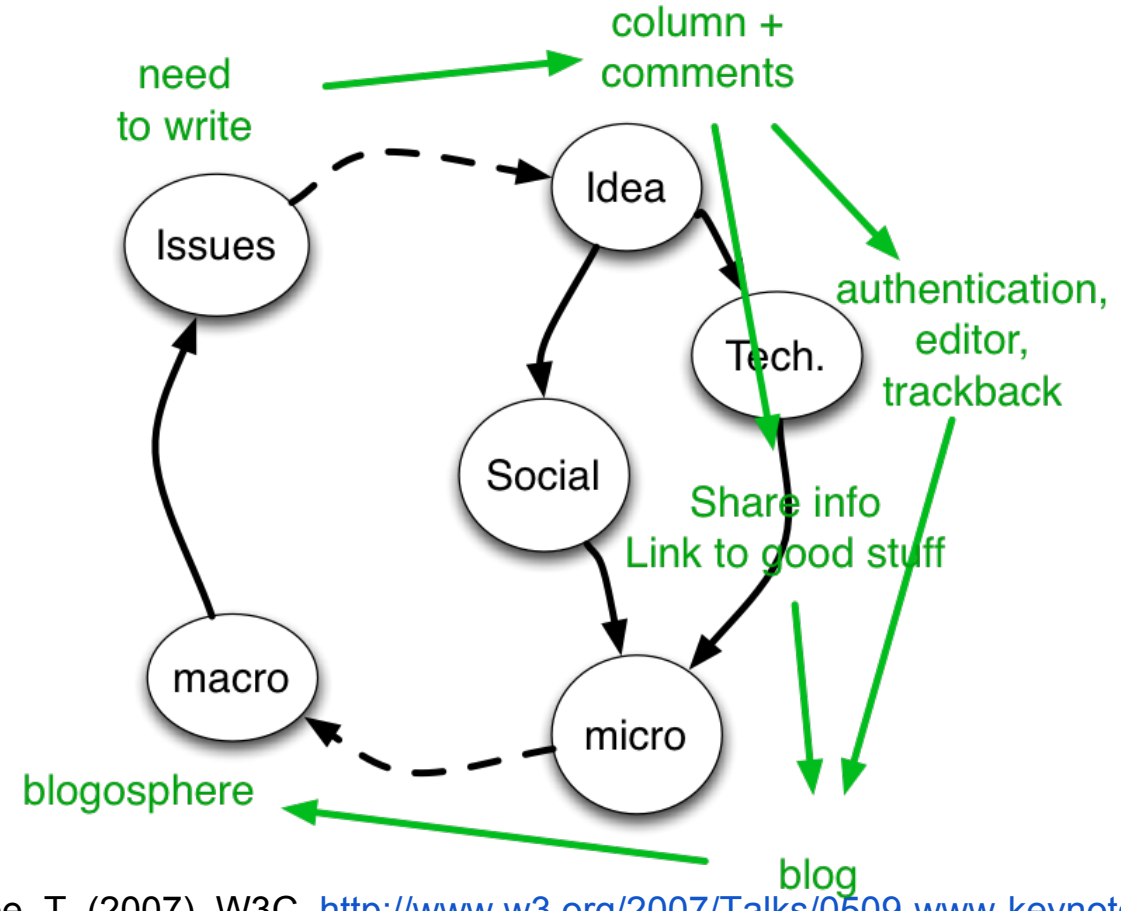
Berners-Lee, T. (2007). W3C. http://www.w3.org/2007/Talks/0509-www-keynote-tbl/#(19)

Applied to Wikis...



Berners-Lee, T. (2007). W3C. http://www.w3.org/2007/Talks/0509-www-keynote-tbl/#(20)

Applied to Blogs...



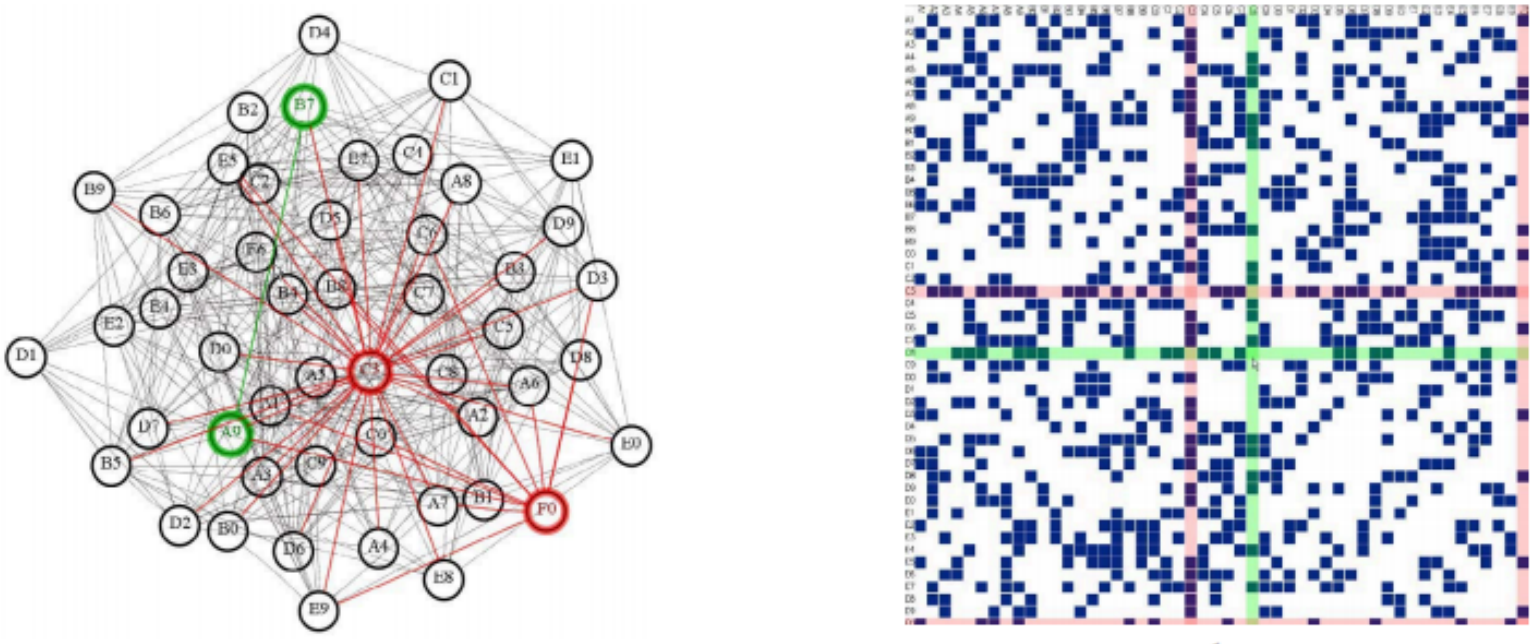
Berners-Lee, T. (2007). W3C. http://www.w3.org/2007/Talks/0509-www-keynote-tbl/#(21)

Social Aspects of the Web

Visual complexity produces opacity. Massive individualizing data produces beautiful, playful hairballs which show us nothing.

- Bruno Latour @ CHI2013

For discussion see, "What baboon notebooks, monads, state surveillance and network diagrams have in common: Bruno Latour at CHI2013" http://bit.ly/14Y3d3u



a *b*

Figure 1 Two visualizations of the same undirected graph containing 50 vertices and 400 edges. The node-link diagram a) is computed using the "neato" program and the matrix representation b) is computed using our VisAdj program.

Multiple disciplines, multiple methods

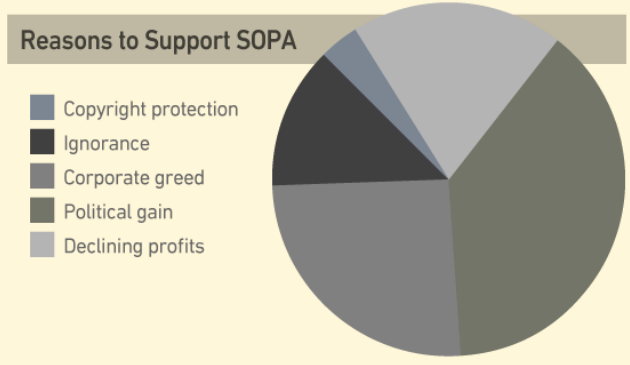
- Given its multiple disciplines, the argument is for a **mixed-methods approach** to measuring the web.
- This means both **quantitative AND qualitative** methods should be employed by researchers.
  - Pros: More robust, comprehensive understanding of “human social behavior”
  - Cons: Diametrically opposed philosophies in data gathering and analysis
- Unanswered questions:
  - Replicability
  - Bias
  - Objectivity and Accuracy
- Ethics

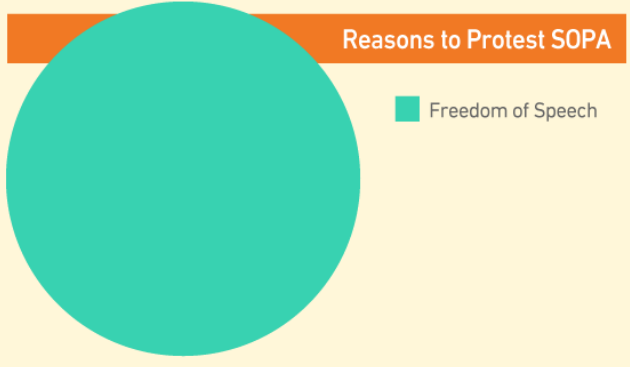
Web Governance, Security and Standards

Who should govern the Web?

W3C
World Wide Web Consortium

SOPA Stop Online Piracy Act





112TH CONGRESS
1ST SESSION

H. R. 3261

To promote prosperity, creativity, entrepreneurship, and innovation by combating the theft of U.S. property, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

OCTOBER 26, 2011

Mr. SMITH of Texas (for himself and Mr. CONYERS, Mr. GOODLATTE, Mr. BERMAN, Mr. GRIFFIN of Arkansas, Mr. GALLEGLY, Mr. DEUTCH, Mr. CHABOT, Mr. ROSS of Florida, Mrs. BLACKBURN, Mrs. BONO MACK, Mr. TERRY, and Mr. SCHIFF) introduced the following bill; which was referred to the Committee on the Judiciary

---

A BILL

To promote prosperity, creativity, entrepreneurship, and innovation by combating the theft of U.S. property, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

(a) SHORT TITLE.—This Act may be cited as the "Stop Online Piracy Act".

(b) TABLE OF CONTENTS.—The table of contents of this Act is as follows:


SUPPORT SOPA
News Corporation
ESPN
MARVEL
viacom
UNIVERSAL
UNIVERSAL MUSIC GROUP
MasterCard
RIAA
VISA
L'ORÉAL
NBCUniversal
MOTION PICTURE ASSOCIATION OF AMERICA
OPPOSE SOPA
Google
Aol.
YAHOO!
facebook
twitter
reddit
zynga
ebay
REPORTERS WITHOUT BORDERS
mozilla
4chan

An important message from

WIKIPEDIA
The Free Encyclopedia

18TH JANUARY, 2012

The Internet needs you

For over a decade, global volunteers have compiled billions of facts and contributed millions of hours to build Wikipedia.

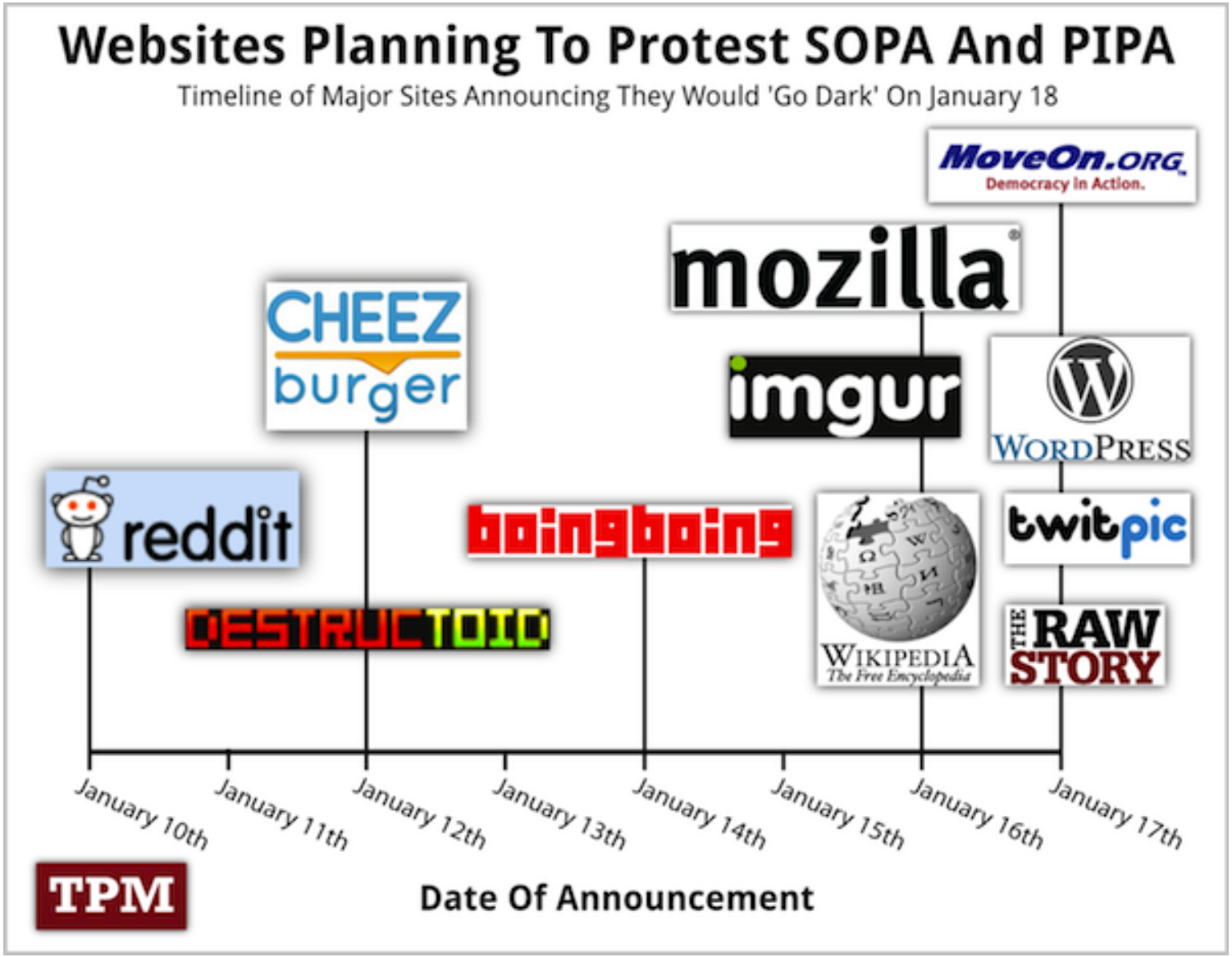
We have only been able to do this because the Internet is free and open; but at this moment, **free speech is in peril like never before**.

The United States Congress is currently considering striking out major rights of free speech and other laws which made Wikipedia possible, forcing us to censor our editor discussions and the information we show you, for the benefit of lobbyists. If passed, it would destroy the freedom of individuals to write without censorship, on every website we have, in any language, everywhere in the world.

Please, consider whether a free and open Internet that includes Wikipedia is something that you, too, care about. Use the tool below to take action. Help protect the Internet. *(more...)*

Here's the call to action

or, continue to Wikipedia


Websites Planning To Protest SOPA And PIPA
Timeline of Major Sites Announcing They Would 'Go Dark' On January 18
MoveOn.org
Democracy in Action.
mozilla
CHEEZ burger
imgur
WordPress
reddit
boingboing
twitpic
DESTRUCTOID
Wikipedia
The Free Encyclopedia
THE RAW STORY
January 10th
January 11th
January 12th
January 13th
January 14th
January 15th
January 16th
January 17th
TPM
Date Of Announcement

Calendar No. 70

112TH CONGRESS
1ST SESSION

S. 968

To prevent online threats to economic creativity and theft of intellectual property, and for other purposes.

IN THE SENATE OF THE UNITED STATES

MAY 12, 2011

Mr. LEAHY (for himself, Mr. HATCH, Mr. GRASSLEY, Mr. SCHUMER, Mrs. FEINSTEIN, Mr. WHITEHOUSE, Mr. GRAHAM, Mr. KOHL, Mr. COONS, Mr. BLUMENTHAL, Ms. KLOBUCHAR, Mr. FRANKEN, Mr. BLUNT, Mr. ALEXANDER, Mrs. GILLIBRAND, and Mr. RUBIO) introduced the following bill; which was read twice and referred to the Committee on the Judiciary

MAY 26, 2011

Reported by Mr. LEAHY, with an amendment

[Strike out all after the enacting clause and insert the part printed in italic]

112TH CONGRESS
1ST SESSION

H. R. 3261

To promote prosperity, creativity, entrepreneurship, and innovation by combating the theft of U.S. property, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

OCTOBER 26, 2011

Mr. SMITH of Texas (for himself and Mr. CONYERS, Mr. GOODLATTE, Mr. BERMAN, Mr. GRIFFIN of Arkansas, Mr. GALLEGLY, Mr. DEUTCH, Mr. CHABOT, Mr. ROSS of Florida, Mrs. BLACKBURN, Mrs. BONO MACK, Mr. TERRY, and Mr. SCHIFF) introduced the following bill; which was referred to the Committee on the Judiciary

A BILL

To promote prosperity, creativity, entrepreneurship, and innovation by combating the theft of U.S. property, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

(a) SHORT TITLE.—This Act may be cited as the "Stop Online Piracy Act".

(b) TABLE OF CONTENTS.—The table of contents of this Act is as follows:

May 12

Sen. Patrick Leahy (D-VT) introduces S. 968, the PROTECT IP Act, a rewrite of the failed Combating Online Infringement and Counterfeits Act of 2010.

Oct. 26

Rep. Lamar Smith (R-TX) introduces H.R. 3261, better known as the Stop Online Piracy Act (SOPA), as the house counterpart to the Senate's PROTECT IP Act.

Nov. 22

The Business Software Alliance, the largest software industry trade group, begins to retreat from its support of SOPA.

Dec. 8

MPAA chief lobbyist, former Senator Chris Dodd, uses China's internet censorship as an example in support of SOPA's site blocking provisions.

Jan. 9

In response to a campaign on popular website Reddit in favor of his opponent, Rep. Paul Ryan (R-WI) announces his opposition to SOPA.

Jan. 13

In the face of public pressure, six GOP Senators ask for the PROTECT IP Act's Jan. 24 vote to be postponed.

Jan. 15

In response to public opposition, the Obama administration announces that it will not support legislation that "...undermines the dynamic, innovative global internet."

Jan. 16

Wikipedia, the largest website so far to do so, announces that it will black out the English language Wikipedia for the entirety of Wednesday, Jan. 18.

Jan. 17

DNS-based site blocking, the most controversial part of SOPA, is removed from both SOPA and PIPA, but the MPAA refuses to renounce future efforts in favor of DNS blocking.

Jan. 18

The great anti-SOPA blackout day. Popular websites ranging from Reddit to Wikipedia are going offline for much of the day, and even Google has joined the protests.

Calendar No. 70

112TH CONGRESS
1ST SESSION

S. 968

To prevent online threats to economic creativity and theft of intellectual property, and for other purposes.

IN THE SENATE OF THE UNITED STATES

MAY 12, 2011

Mr. LEAHY (for himself, Mr. HATCH, Mr. GRASSLEY, Mr. SCHUMER, Mrs. FEINSTEIN, Mr. WHITEHOUSE, Mr. GRAHAM, Mr. KOHL, Mr. COONS, Mr. BLUMENTHAL, Ms. KLOBUCHAR, Mr. FRANKEN, Mr. BLUNT, Mr. ALEXANDER, Mrs. GILLIBRAND, and Mr. RUBIO) introduced the following bill; which was read twice and referred to the Committee on the Judiciary

MAY 26, 2011

Reported by Mr. LEAHY, with an amendment

[Strike out all after the enacting clause and insert the part printed in italic]

112TH CONGRESS
1ST SESSION

H. R. 3261

To promote prosperity, creativity, entrepreneurship, and innovation by combating the theft of U.S. property, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

OCTOBER 26, 2011

Mr. SMITH of Texas (for himself and Mr. CONYERS, Mr. GOODLATTE, Mr. BERMAN, Mr. GRIFFIN of Arkansas, Mr. GALLEGLY, Mr. DEUTCH, Mr. CHABOT, Mr. ROSS of Florida, Mrs. BLACKBURN, Mrs. BONO MACK, Mr. TERRY, and Mr. SCHIFF) introduced the following bill; which was referred to the Committee on the Judiciary

A BILL

To promote prosperity, creativity, entrepreneurship, and innovation by combating the theft of U.S. property, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

(a) SHORT TITLE.—This Act may be cited as the "Stop Online Piracy Act".

(b) TABLE OF CONTENTS.—The table of contents of this Act is as follows:

Top 5 Industries, 2013-2014, Campaign Cmte and Leadership PAC

Industry	Total	Indivs	PACs
TV/Movies/Music	$25,700	$6,200	$19,500
Computers/Internet	$25,460	$250	$25,210
Retired	$23,050	$23,050	$0
Oil & Gas	$22,250	$250	$22,000
Insurance	$19,750	$8,750	$11,000

Top 5 Industries, 2009-2014, Campaign Cmte and Leadership PAC

Industry	Total	Indivs	PACs
Lawyers/Law Firms	$674,691	$516,781	$157,910
TV/Movies/Music	$535,506	$310,856	$224,650
Lobbyists	$408,050	$396,550	$11,500
Computers/Internet	$200,620	$105,950	$94,670
Leadership PACs	$147,400	$0	$147,400

May 12

Sen. Patrick Leahy (D-[illegible] introduces S. 968, the PROTECT IP Act, a rewrite of the failed Combating Online Infringement and Counterfeits Act of 20[illegible]

Oct. 26

Rep. Lamar Smith (R-TX) introduces H.R. 3261, better [illegible]n as the Stop Online [illegible]'SOPA), as the [illegible]rt to the [illegible]mines the dynamic, innovative global internet."

Nov. 22

The Business Software Alliance, the largest software industry trade group, begins to retreat from its support of SOPA.

[illegible] entirety of Wednesday, Jan. 18.

Dec. 8

MPAA chief lobbyist, former Senator Chris Dodd, uses China's internet censorship as an example in support of SOPA's site blocking provisions.

Jan. 9

In response to a campaign on popular website Reddit in favor of his opponent, Rep. Paul Ryan (R-WI) announces his opposition to SOPA.

Jan. 13

In the face of public pressure, six GOP Sen[illegible] tors ask for the PROT[illegible] IP Act's Jan. 24 vote t[illegible] postponed.

Jan. 17

DNS-based site blocking, the most controversial part of SOPA, is removed from both SOPA and PIPA, but the MPAA refuses to renounce future efforts in favor of DNS blocking.

Jan. 18

The great anti-SOPA blackout day. Popular websites ranging from Reddit to Wikipedia are going offline for much of the day, and even Google has joined the protests.

Calendar No. 70

112TH CONGRESS
1ST SESSION

S. 968

To prevent online threats to economic creativity and theft of intellectual property, and for other purposes.

IN THE SENATE OF THE UNITED STATES

MAY 12, 2011

Mr. LEAHY (for himself, Mr. HATCH, Mr. GRASSLEY, Mr. SCHUMER, Mrs. FEINSTEIN, Mr. WHITEHOUSE, Mr. GRAHAM, Mr. KOHL, Mr. COONS, Mr. BLUMENTHAL, Ms. KLOBUCHAR, Mr. FRANKEN, Mr. BLUNT, Mr. ALEXANDER, Mrs. GILLIBRAND, and Mr. RUBIO) introduced the following bill; which was read twice and referred to the Committee on the Judiciary

MAY 26, 2011

Reported by Mr. LEAHY, with an amendment

[Strike out all after the enacting clause and insert the part printed in italic]

112TH CONGRESS
1ST SESSION

H. R. 3261

To promote prosperity, creativity, entrepreneurship, and innovation by combating the theft of U.S. property, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

OCTOBER 26, 2011

Mr. SMITH of Texas (for himself and Mr. CONYERS, Mr. GOODLATTE, Mr. BERMAN, Mr. GRIFFIN of Arkansas, Mr. GALLEGLY, Mr. DEUTCH, Mr. CHABOT, Mr. ROSS of Florida, Mrs. BLACKBURN, Mrs. BONO MACK, Mr. TERRY, and Mr. SCHIFF) introduced the following bill; which was referred to the Committee on the Judiciary

A BILL

To promote prosperity, creativity, entrepreneurship, and innovation by combating the theft of U.S. property, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

(a) SHORT TITLE.—This Act may be cited as the "Stop Online Piracy Act".

(b) TABLE OF CONTENTS.—The table of contents of this Act is as follows:

Top 5 Industries, 2013-2014, Campaign Cmte and Leadership PAC

Industry	Total	Indivs	PACs
TV/Movies/Music	$25,700	$6,200	$19,500
Computers/Internet	$25,460	$250	$25,210
Retired	$23,050	$23,050	$0
Oil & Gas	$22,250	$250	$22,000
Insurance	$19,750	$8,750	$11,000

May 12

Sen. Patrick Leahy (D-[illegible] introduces S. 968, the PROTECT IP Act, a rewrite of the failed Combating Online Infringement and Counterfeits Act of 20[illegible]

Oct. 26

Rep. Lamar Smith (R-TX) introduces H.R. 3261, better [illegible]n as the Stop Online [illegible]'SOPA), as the [illegible]rt to the [illegible]

Nov. 22

The Business Software Alliance, the largest software industry trade group, begins to retreat from its support of SOPA.

Dec. 8

MPAA chief lobbyist, former Senator Chris Dodd, uses China's internet cen[illegible]sor[illegible] sup[illegible] blo[illegible]

Jan. 9

In response to a campaign on popular website Reddit in favor of his opponent,

Jan. 13

In the face of public pressure, six GOP Sen[illegible]tors ask for the PROT[illegible] IP Act's Jan. 24 vote t[illegible] postponed.

Top 5 Industries, 2009-2014, Campaign Cmte and Leadership PAC

Industry	Total	Indivs	PACs
Lawyers/Law Firms	$674,691	$516,781	$157,910
TV/Movies/Music	$535,506	$310,856	$224,650
Lobbyists	$408,050	$396,550	$11,500
Computers/Internet	$200,620	$105,950	$94,670
Leadership PACs	$147,400	$0	$147,400

[illegible]mines the dynamic, innovative global internet."

[illegible]entirety of Wednesday, Jan. 18.

Ja

DN[illegible] the[illegible] par[illegible] fro[illegible] but[illegible] ren[illegible] fav[illegible]

Members of Congress's Positions on SOPA/PIPA, as tracked by ProPublica.org

Jan. 18

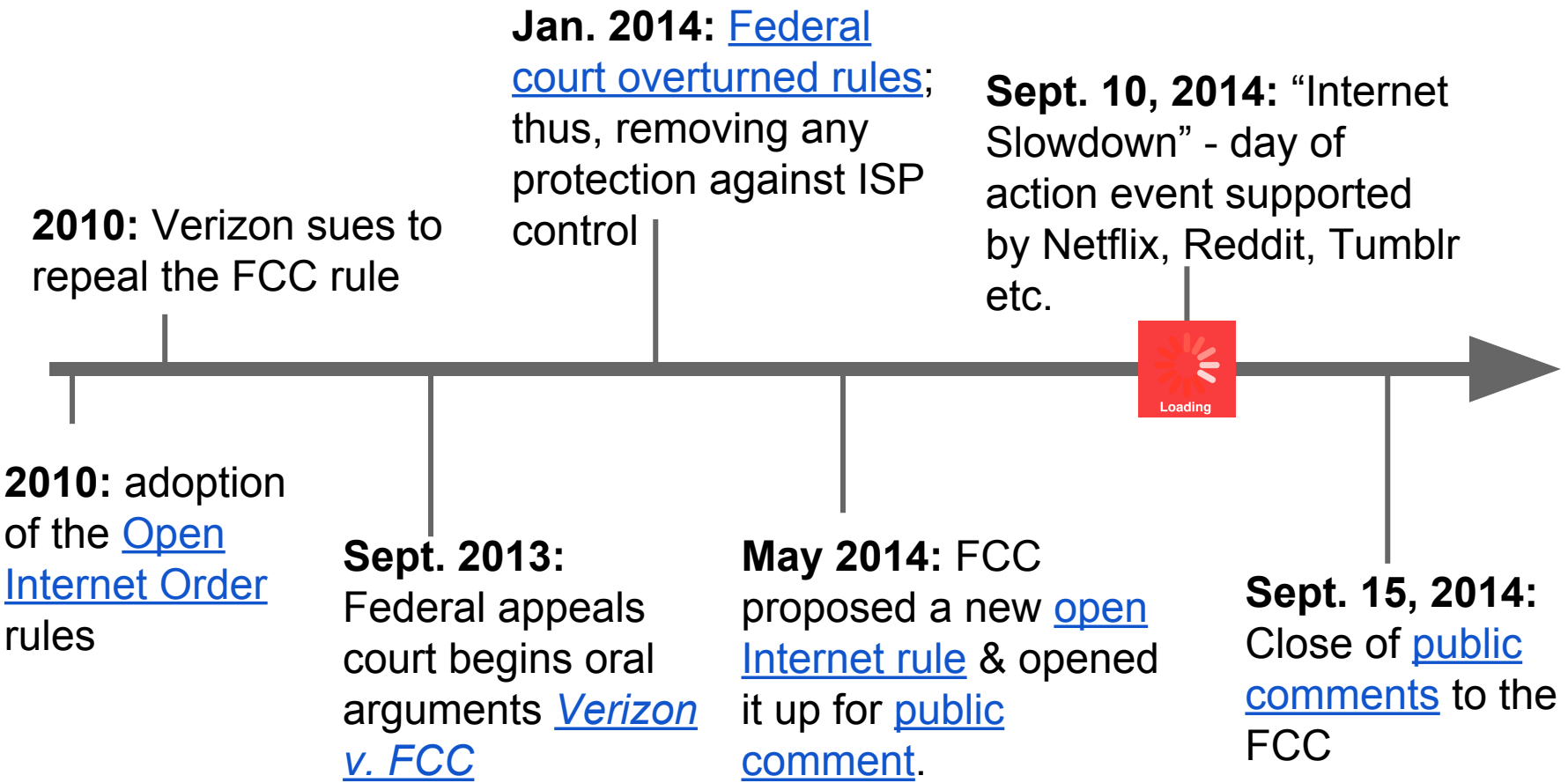
80 supporters 31 opponents

Jan. 19

65 supporters 101 opponents


Loading

Net Neutrality (simplified)



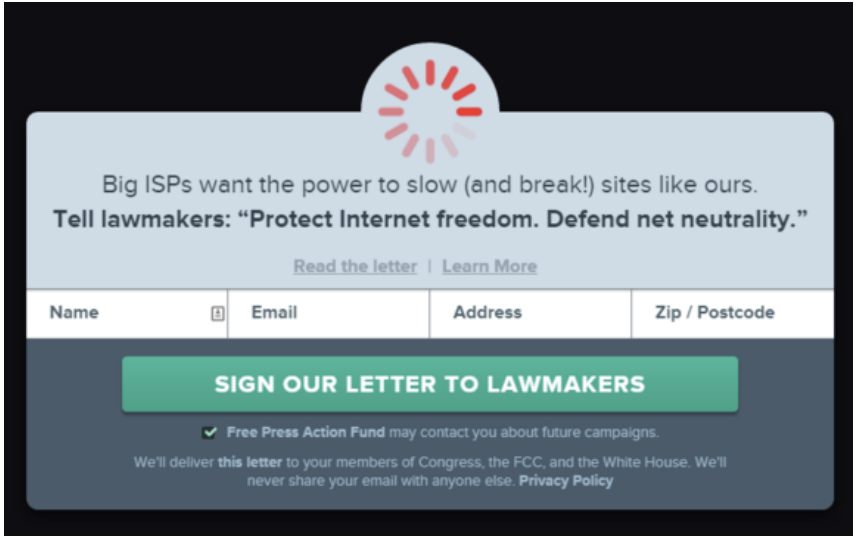
PROS

- Equal access to the Internet
- Preservation of free speech
- Prevents further fragmentation of the Internet

AGAINST

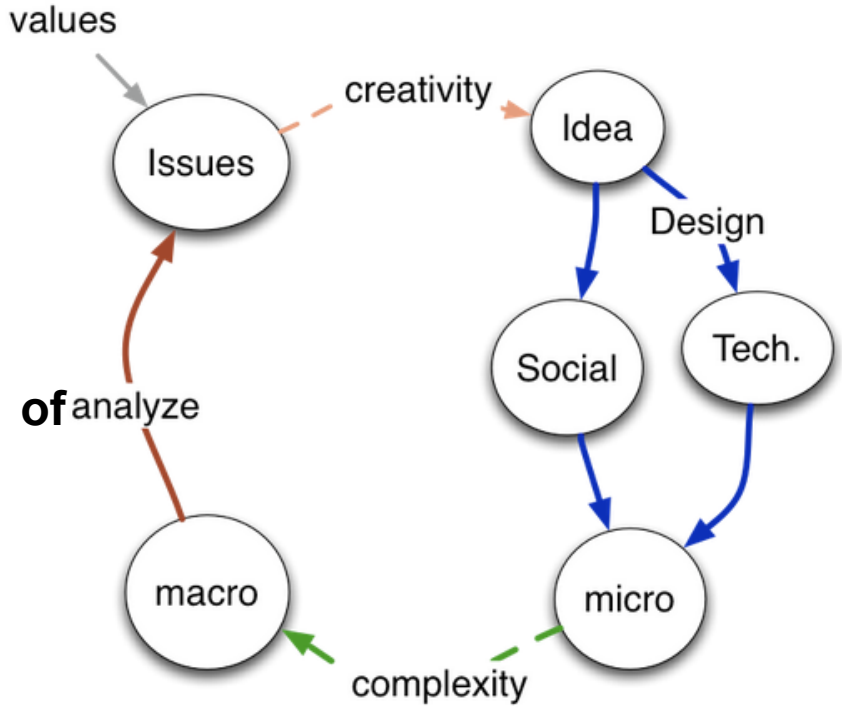
- Preservation of open commerce
- Pay for better quality of service (fast lane)
- Increase of competition

Internet Slow Down: Response by the numbers



- **Over 10,000 sites** participated in this day of action. Including, Netflix, reddit & Tumblr.
- **+ 111,000 new comments** were registered with the FCC (largest amount since the Janet Jackson wardrobe malfunction)
- **1.5 million** emails were sent to Congress
- **+ 286,000** calls were made to Congress (avg. 1,000 calls per hour)

Web Science meets (Web) Governance?



Policy Conception

Policy Design

Policy Implementation

Analysis and understanding of policy implications

Review...

1. A Science of The Web
2. Web Architecture
3. Measuring the Web
4. The Web Science Method
5. Social Aspects of the Web
6. Web and other Governance

Assignment:

1. NetLogo "Preferential Attachment" Simulator: http://bit.ly/18bd0p2
   - Try the THINGS TO TRY!
2. Excel-based Network Analysis Tutorial
   - Following instructions at: http://bit.ly/1a3mtzW
   - Install **NodeXL** from: http://bit.ly/1qYOd3N
   - Use **Senate 2007** data from: http://bit.ly/1a3mhAI
   - Play with other data at: http://bit.ly/1a3mJ1X
3. Social Network Exploration
   - Twitalizer: http://twitalyzer.com
   - TweetArchivist: http://tweetarchivist.com
   - MentionMap: http://mentionmapp.com
4. Create a Web Science Scenario:
   - Identify a (social) problem
   - Propose an engineered solution
   - Identify how to measure, analyze, evaluate, iterate...