CSS2013 Beihang

Computational Social Science: A Brief Intro

The CSS lecturers:
Lexing Xie, Rob Ackland, Paul Thomas, Jamsheed Shorish

Lecture slides credit: Lada Adamic, Univ. Michigan
Jure Leskovec, Stanford University

Computational Social Science: the coming of age

"I realized that computer science is not just about technology, It is also a human topic."
-- Jon Kleinberg

"It's been really transformative, … We were limited before to surveys, which are retrospective, and lab experiments, which are almost always done on small numbers of college sophomores. I liken the opportunities to the changes in physics brought about by the particle accelerator, and in neuroscience by functional magnetic resonance imaging".
-- Michael Macy

NATURE | NEWS FEATURE

Computational social science: Making the links

From e-mails to social networks, the digital traces left by life in the modern world are transforming social science.

Jim Giles

22 August 2012 | Corrected: 31 August 2012

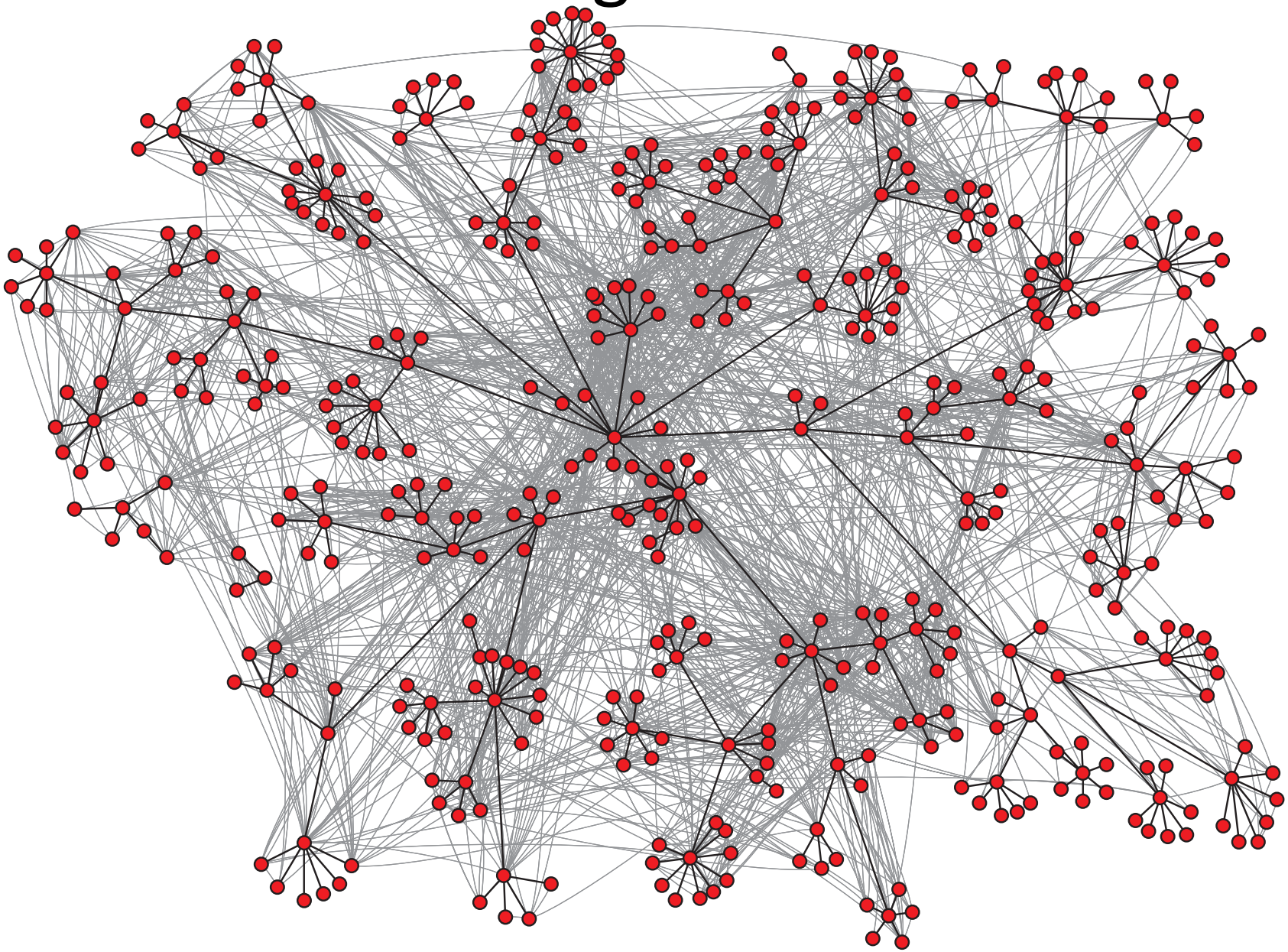
Jon Kleinberg's early work was not for the mathematically faint of heart. His first publication[1], in 1992, was a computer-science paper with contents as dense as its title: 'On dynamic Voronoi diagrams and the minimum Hausdorff distance for point sets in the plane'.

http://www.nature.com/news/computational-social-science-making-the-links-1.11243

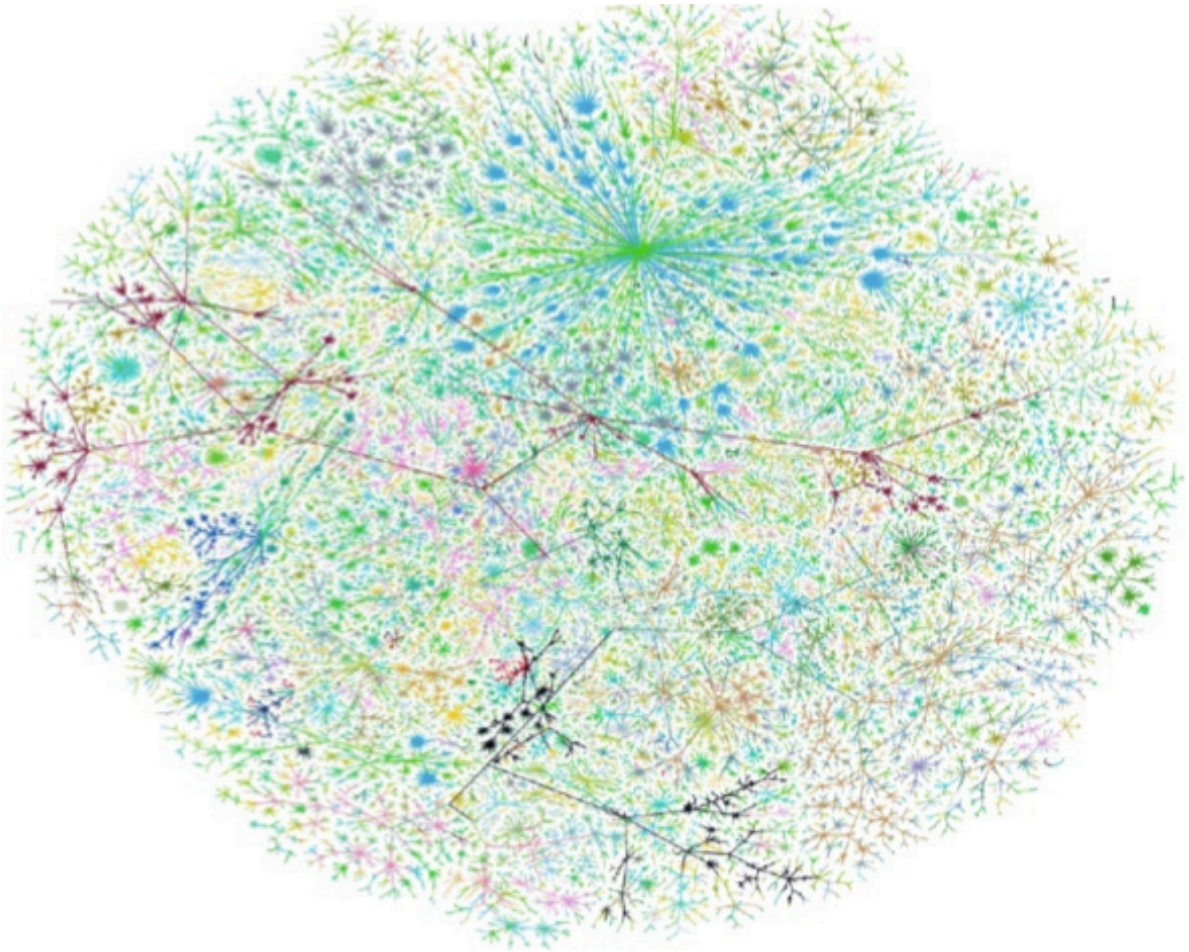
A View of Facebook via 10 M links

By Paul Butler, https://www.facebook.com/note.php?note_id=469716398919

Networks: An organization - HP Labs

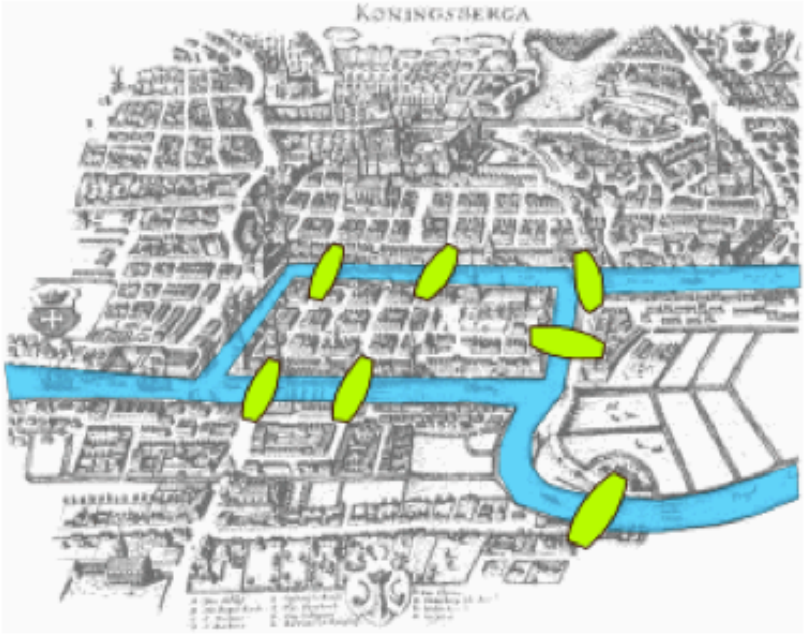


Networks: Communications



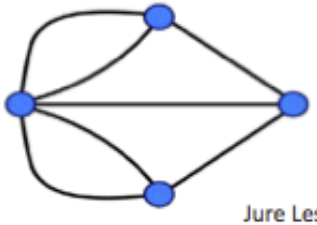
Graph of the Internet
(Autonomous Systems)

Networks: Transportation



Seven Bridges of Königsberg (Euler 1735)

Return to the starting point by traveling each link of the graph once and only once.

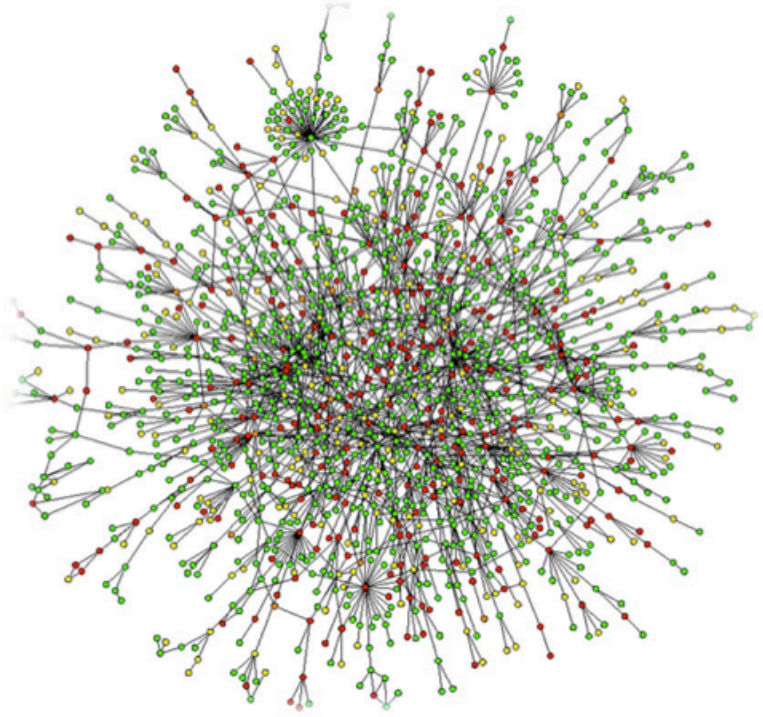


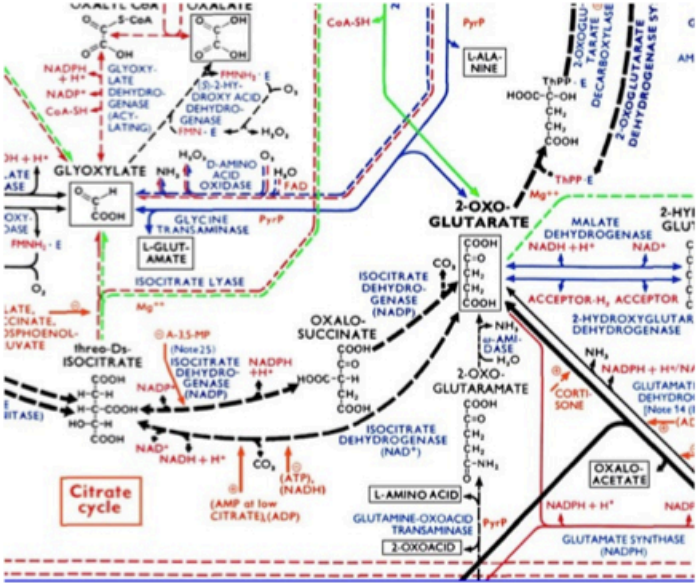
London Underground

Networks: Brain

Human brain has between 10-100 billion neurons

Networsk: Cells

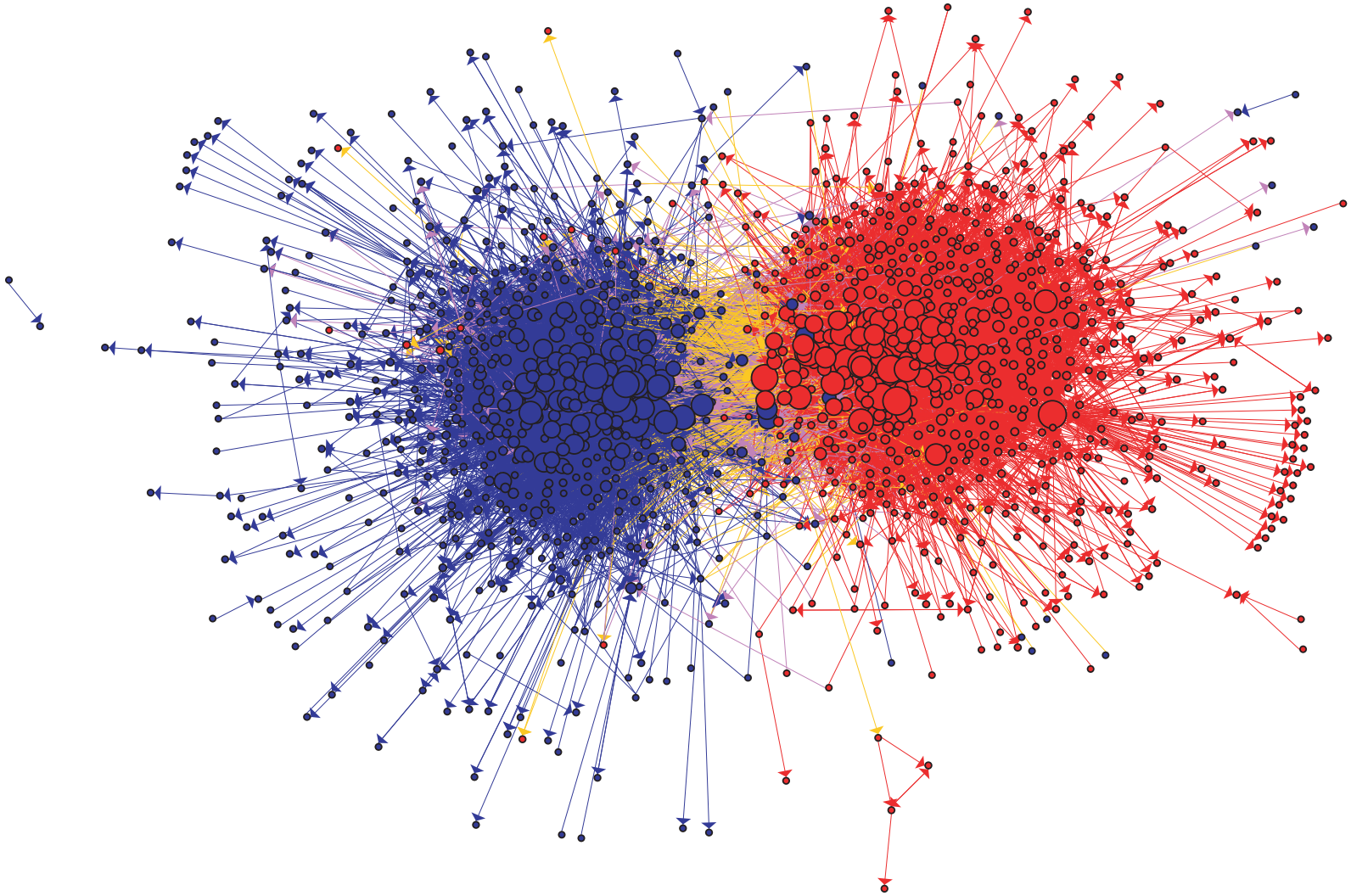




Protein-Protein Interaction Networks:
Nodes: Proteins
Edges: 'physical' interactoins

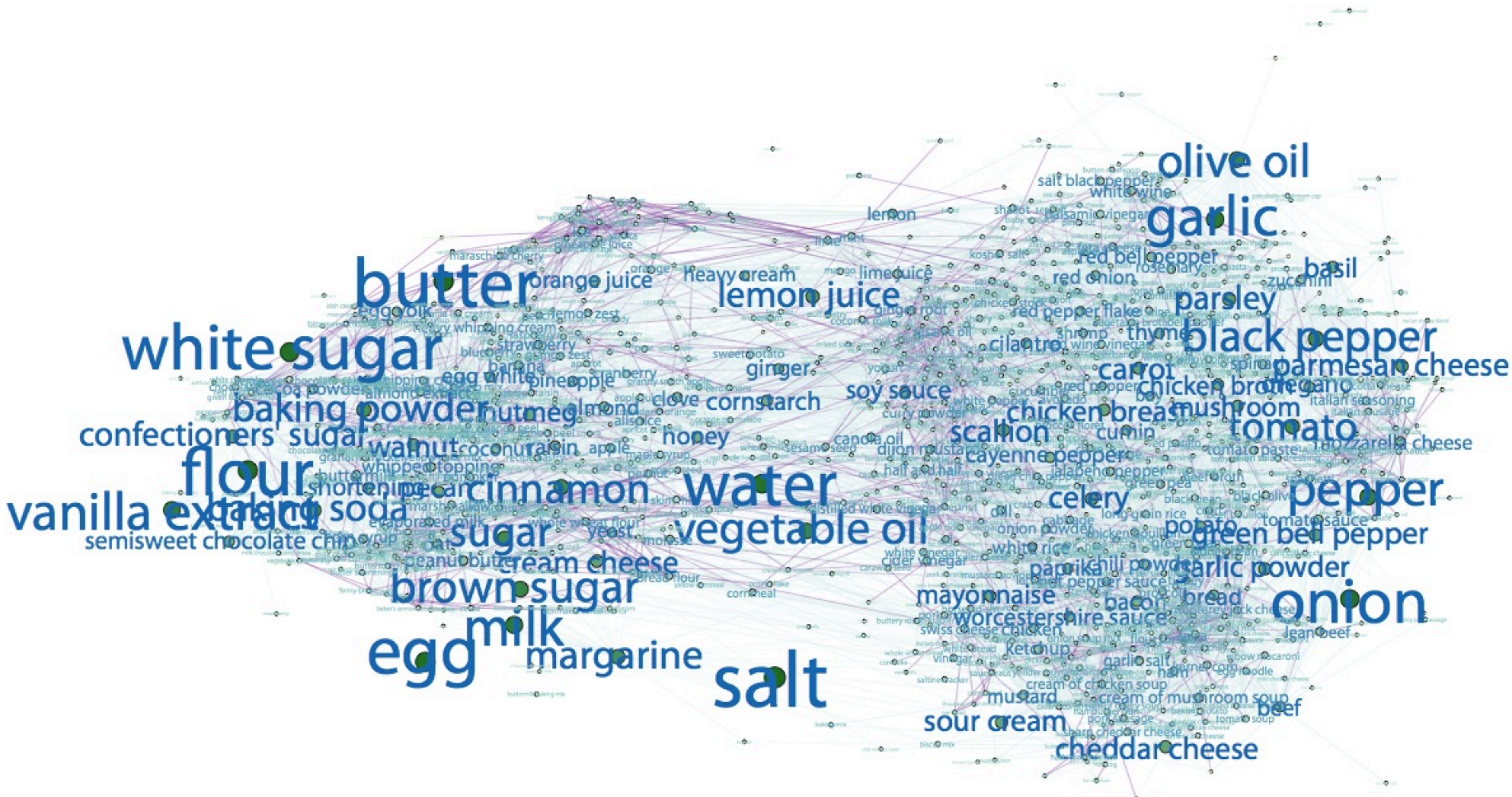
Metabolic networks:
Nodes: Metabolites and enzymes
Edges: Chemical reactions

Political blogs



Lada A. Adamic and Natalie Glance. 2005. The political blogosphere and the 2004 U.S. election: divided they blog. In Proceedings of the 3rd international workshop on Link discovery (LinkKDD '05)

Ingredient networks



Recipe recommendation using ingredient networks. Chun-Yuen Teng, Yu-Ru Lin, Lada A. Adamic, WebSci 2011.

Why Networks?

- Behind each of these complex systems there is an intricate wiring diagram,
  a network
  that defines the interactions between system components.

Understanding the network is key to understand the behaviors of such complex system.

Application domains in network analysis

This class

- Social (people-people) networks ✔
- Information networks ✔
- Organization and political networks ✔

- Computer networking
- Biology
- Transportation networks

CSS: Computational challenges

This class

- Machine Learning and applied statistics:
  Predictive modeling, models for network structure, models for dynamic events, optimization ✔
- Natural Language Processing:
  linguistic styles and variations, document similarity, information retrieval ✔
- Visual analytics, visualization ✔
- Computer systems: scalability, reliability, programming languages and tools ✗
- …

Database and data warehouse 1.0

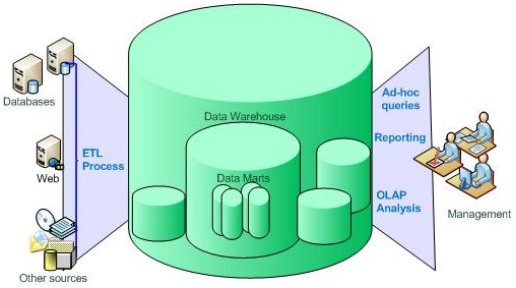
Computer + programming interfaces

Database

Database query language

SQL

Data warehouse



Visualizing Friendships

by Paul Butler for Facebook Engineering (Notes) on Tuesday, December 14, 2010 at 10:16am

Visualizing data is like photography. Instead of starting with a blank canvas, you manipulate the lens used to present the data from a certain angle.

1. Sample 10 million links from Hive
2. Get city/lat/lon for each node
3. Compute connection strength between cities
4. Render links among city pairs in R
5. Tweak rendering and path drawing until satisfactory

https://www.facebook.com/note.php?note_id=469716398919

Storing and Manipulating 500M Links

"HUGE" Computer + programming interfaces

APACHE HADOOP

hadoop

The Apache Hadoop project develops open-source software for reliable, scalable, distributed computing.

Database

APACHE HBASE

Data-centric platform and query language interface

Data warehouse with SQL-like interface

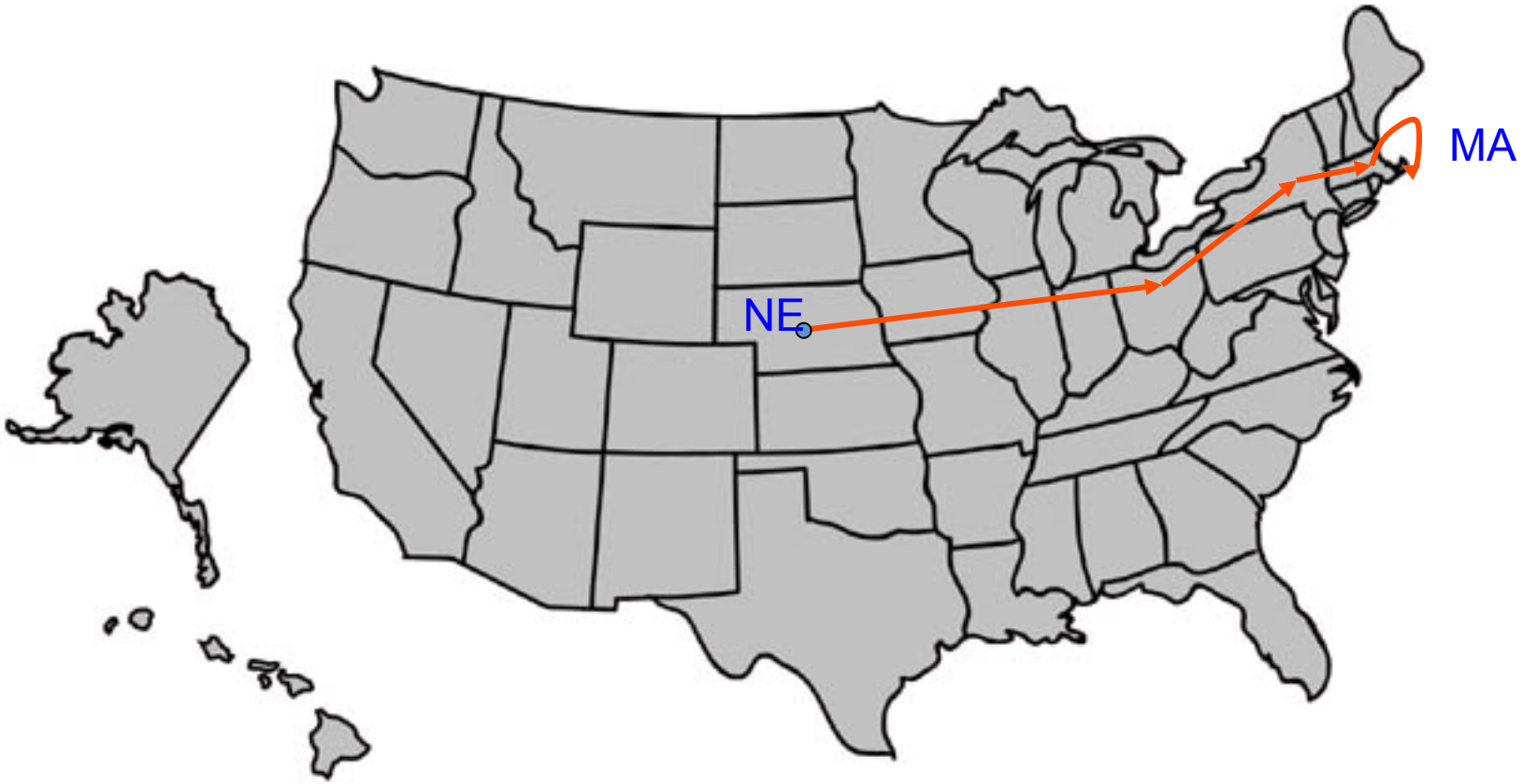
HIVE

http://developer.yahoo.com/blogs/hadoop/posts/2010/08/pig_and_hive_at_yahoo/

What are the classic, solve, and open questions in CSS?

PROBLEMS IN CSS

Small-world phenomena: Milgram's Experiment



http://en.wikipedia.org/wiki/Small_world_experiment 1967 -- 1969

Slide by Lada Adamic, U Michigan

Milgram's experiment

Instructions:
Given a target individual (stockbroker in Boston), pass the message to a person you correspond with who is "closest" to the target.

Outcome:

20% of initiated chains reached target
average chain length = 6.5

- "Six degrees of separation"

Milgram's experiment repeated

email experiment
Dodds, Muhamad, Watts,
Science 301, (2003)
(optional reading)

•18 targets
•13 different countries

•60,000+ participants
•24,163 message chains
•384 reached their targets
•average path length 4.0

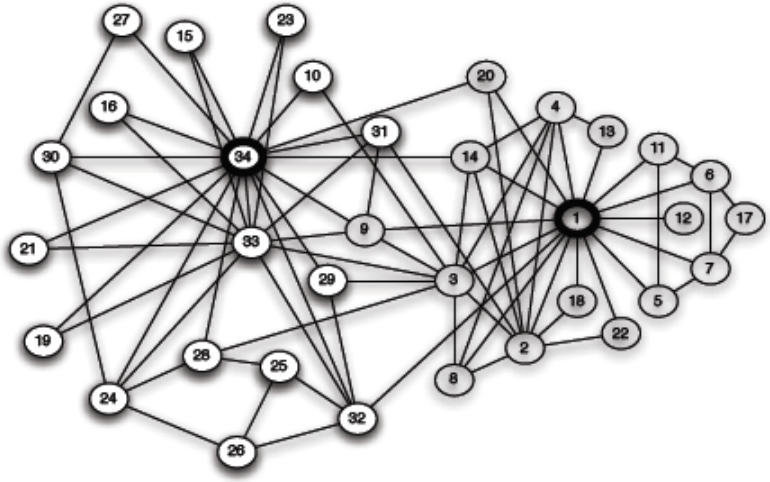
Source: NASA, U.S. Government; http://visibleearth.nasa.gov/view_rec.php?id=2429

Slide by Lada Adamic, U Michigan

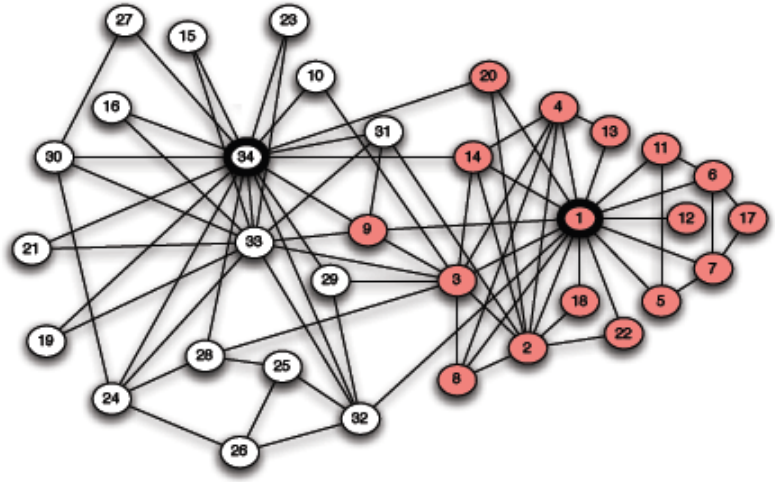
- Is 6 is a ***surprising*** number?
  In the 1960s? Today? Why?

  What is the mechanism behind “small-world” networks?

Zachary Karate Club



(a) *Karate club network*



(b) *After a split into two clubs*

source:Easley/Kleinberg

Slide by Lada Adamic, U Michigan

Community detection

- Why detect communities?
  - Simplified network structure and “big picture”
  - Explain actions and positons
  - Better predictions
- What defines a community?
  - Interaction
  - Profile
  - Dynamics
  - …

Link prediction

"Given a snapshot of a social network, can we infer which new interactions among its members are likely to occur in the near future? "

What are the measurements?

-- "proximity" and "similarity" between two unconnected nodes

What are application domains?

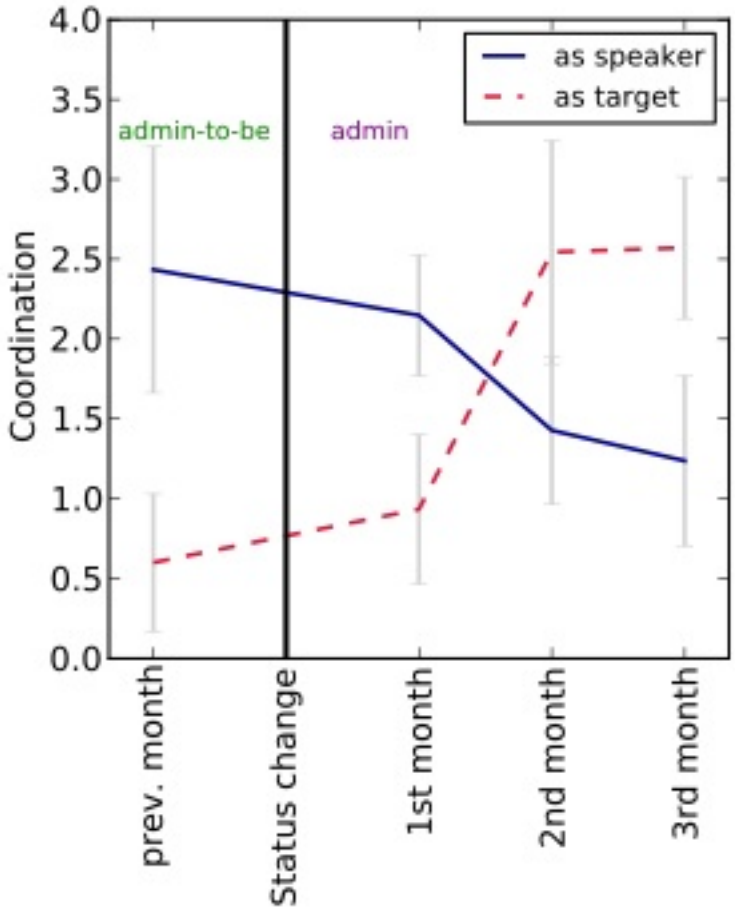
-- social networks, friend recommendation; product / webpage recommendation; predicting academic collaborations; predicting merger and acquisitions …

How to measure performance?

-- use future held-out data, conversion rate, …

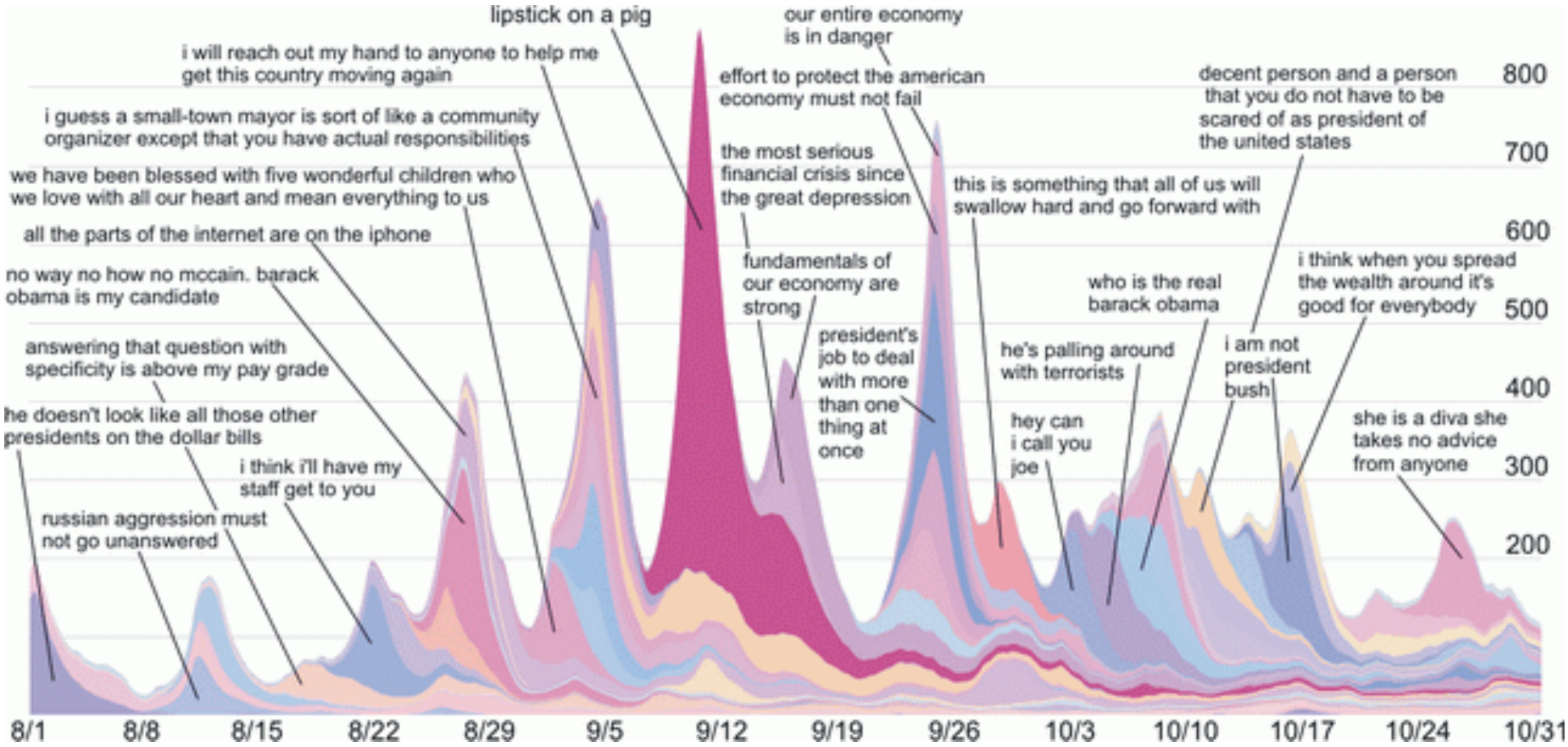
Liben-Nowell, D. and Kleinberg, J. The link-prediction problem for social networks. Journal of the American Society for Information Science and Technology, 58(7) 1019{1031 (2007)

Language Use in Social Media



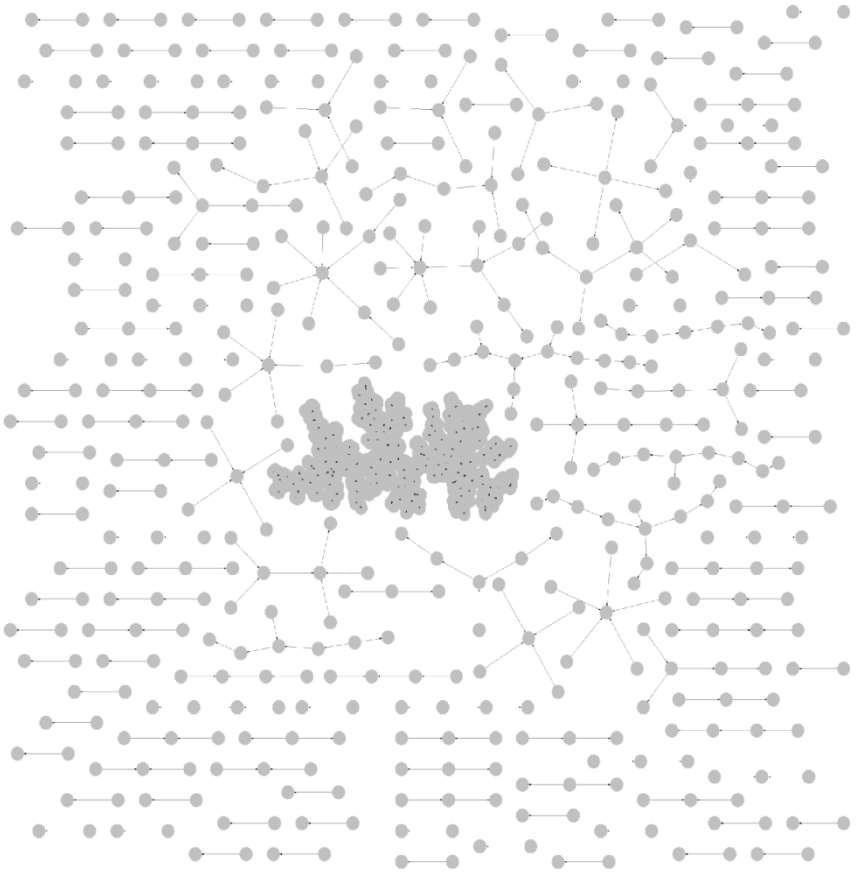
Echoes of power: Language effects and power differences in social interaction
Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang and Jon Kleinberg.
Proceedings of WWW, 2012.

Tracking Memes

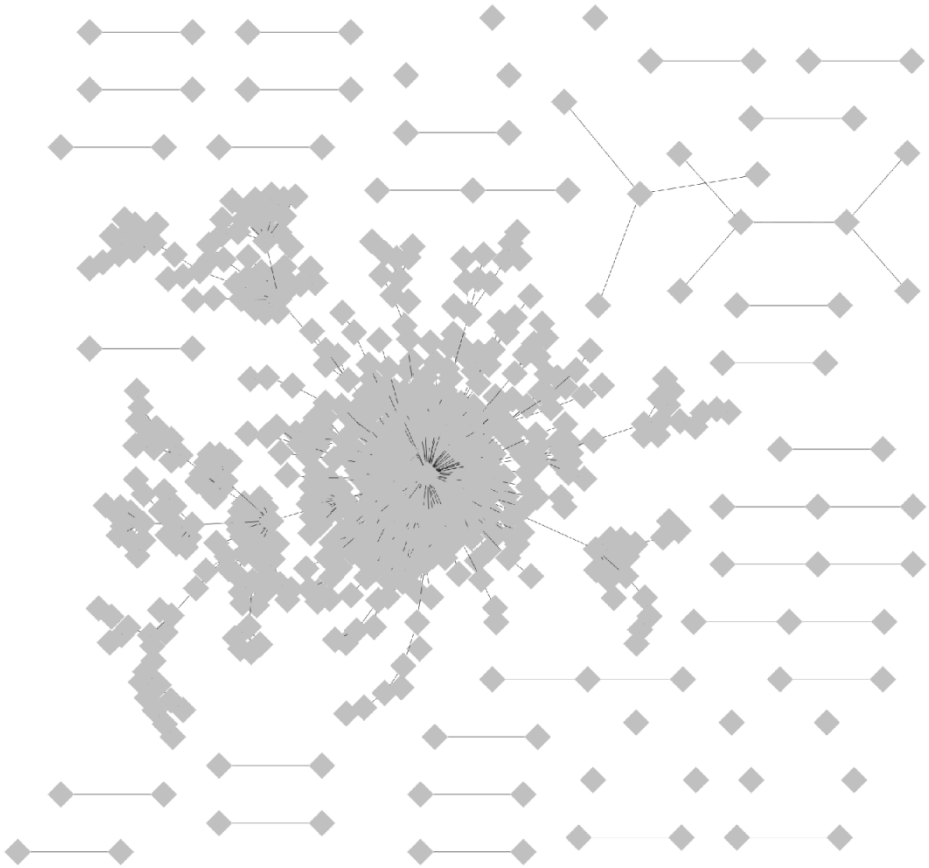


YouTube Remix Network

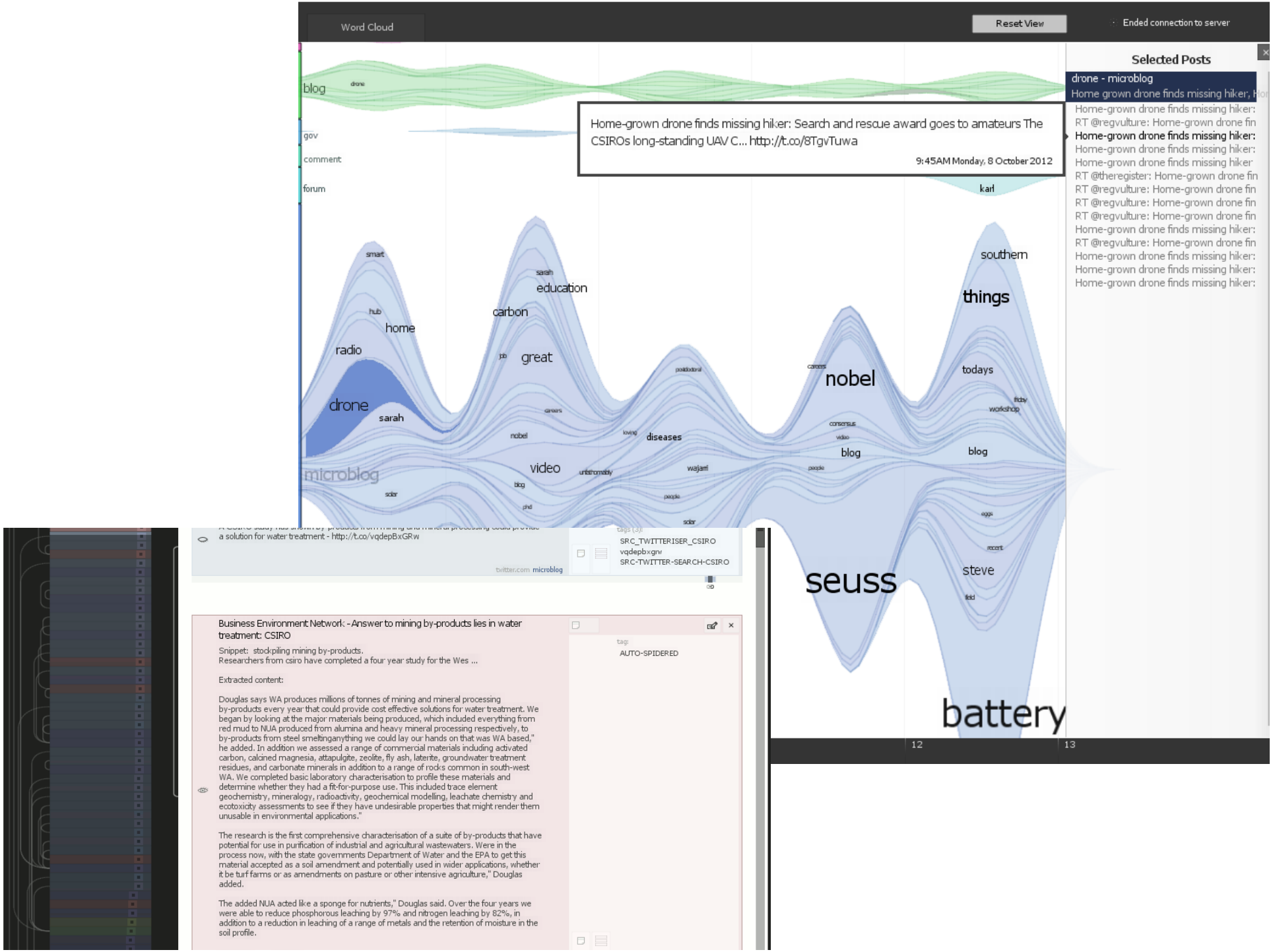
Video graph



Author graph



Visual Memes in Social Media: Tracking Real-world News in YouTube Videos (2011), Lexing Xie, Apostol Natsev, Matthew Hill, John Kender, John R Smith, ACM Multimedia 2011, Scottsdale, AZ, USA, Nov 2011


Word Cloud
Reset View
Ended connection to server
Selected Posts
drone - microblog
Home grown drone finds missing hiker, H
Home-grown drone finds missing hiker:
RT @regvulture: Home-grown drone fin
Home-grown drone finds missing hiker:
Home-grown drone finds missing hiker:
Home-grown drone finds missing hiker
RT @theregister: Home-grown drone fin
RT @regvulture: Home-grown drone fin
RT @regvulture: Home-grown drone fin
RT @regvulture: Home-grown drone fin
Home-grown drone finds missing hiker:
RT @regvulture: Home-grown drone fin
Home-grown drone finds missing hiker:
Home-grown drone finds missing hiker:
Home-grown drone finds missing hiker:
Home-grown drone finds missing hiker: Search and rescue award goes to amateurs The CSIROs long-standing UAV C... http://t.co/8TgvTuwa
9:45AM Monday, 8 October 2012
blog
gov
comment
forum
drone
karl
smart
sarah
education
hub
carbon
home
radio
great
drone
sarah
nobel
diseases
video
microblog
southern
things
nobel
todays
workshop
blog
blog
seuss
steve
battery
12
13
a solution for water treatment - http://t.co/vqdepBxGRw
twitter.com microblog
tags (3):
SRC_TWITTERISER_CSIRO
vqdepbxgrw
SRC-TWITTER-SEARCH-CSIRO
Business Environment Network - Answer to mining by-products lies in water treatment: CSIRO
Snippet: stockpiling mining by-products.
Researchers from csiro have completed a four year study for the Wes ...
Extracted content:
Douglas says WA produces millions of tonnes of mining and mineral processing by-products every year that could provide cost effective solutions for water treatment. We began by looking at the major materials being produced, which included everything from red mud to NUA produced from alumina and heavy mineral processing respectively, to by-products from steel smeltinganything we could lay our hands on that was WA based," he added. In addition we assessed a range of commercial materials including activated carbon, calcined magnesia, attapulgite, zeolite, fly ash, laterite, groundwater treatment residues, and carbonate minerals in addition to a range of rocks common in south-west WA. We completed basic laboratory characterisation to profile these materials and determine whether they had a fit-for-purpose use. This included trace element geochemistry, mineralogy, radioactivity, geochemical modelling, leachate chemistry and ecotoxicity assessments to see if they have undesirable properties that might render them unusable in environmental applications."
The research is the first comprehensive characterisation of a suite of by-products that have potential for use in purification of industrial and agricultural wastewaters. Were in the process now, with the state governments Department of Water and the EPA to get this material accepted as a soil amendment and potentially used in wider applications, whether it be turf farms or as amendments on pasture or other intensive agriculture," Douglas added.
The added NUA acted like a sponge for nutrients," Douglas said. Over the four years we were able to reduce phosphorous leaching by 97% and nitrogen leaching by 82%, in addition to a reduction in leaching of a range of metals and the retention of moisture in the soil profile.
tag:
AUTO-SPIDERED

08:19 AM #1

Caption

er
32

Join Date: Sep 2010
Posts: 9,209
Thanks: 1,961
Thanked 1,611 Times i
Mentioned: 0 Post(s)
Tagged: 0 Thread(s)
Rep Power: 369

fuse To Cover Kansas Schools Where Teachers Carry Guns: It's Too Risky

thrusts Iowa insurer into gun debate

MC Insurance Cos. insures 85 percent to 90 percent of all Kansas school districts and has refused to renew coverage for schools that permit teachers and custodians to carry concealed ns on their campuses under the new law, which took effect July 1. It's not a political decision, but a financial one based on the riskier climate it estimates would be created, the insure

been writing school business for almost 40 years, **and one of the underwriting guidelines we follow for schools is that any on-site armed security should be provided by uniformed, qualifi ement officers,"** said Mick Lovell, EMC's vice president for business development. "Our guidelines have not recently changed."

sly hate America...or something

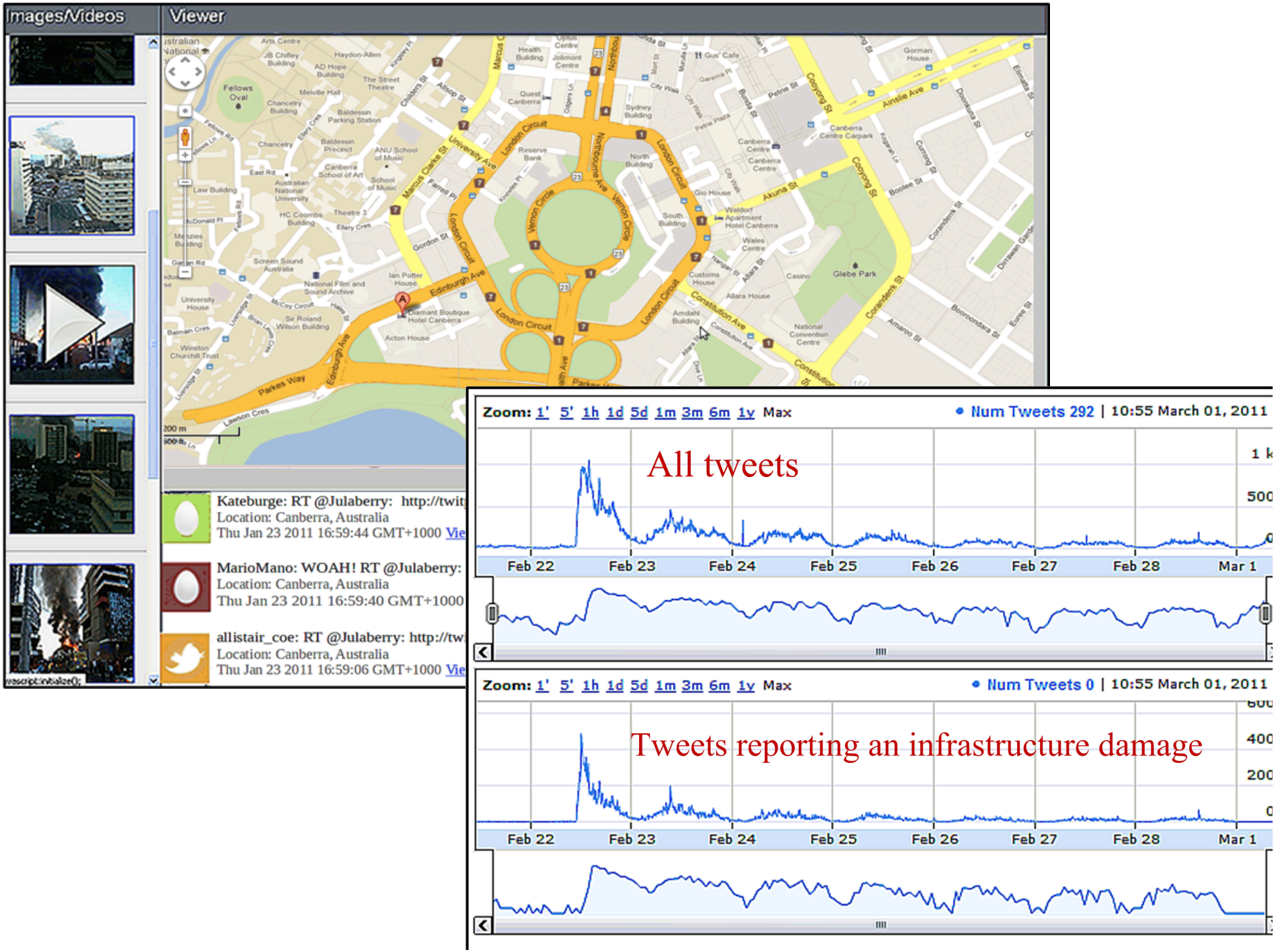
nce is all about risk and about pricing the cost of coverage in a way that correctly reflects it. That's one of the reasons many schools have gotten rid of their trampolines, he said.

e thing to have a trained peace officer with a gun in school; it's a completely different situation when you have a custodian or a teacher with a gun," Skow said. "That changes the risk of insuring a scho ies it considerably."

a trampoline can hurt one person, modern weapons have the potential to kill many people very quickly, he said

vent downhill quickly

Originally Posted by TemplarKormac


Images/Videos
Viewer
Kateburge: RT @Julaberry: http://twit
Location: Canberra, Australia
Thu Jan 23 2011 16:59:44 GMT+1000 Vie
MarioMano: WOAH! RT @Julaberry:
Location: Canberra, Australia
Thu Jan 23 2011 16:59:40 GMT+1000
allistair_coe: RT @Julaberry: http://tw
Location: Canberra, Australia
Thu Jan 23 2011 16:59:06 GMT+1000 Vie
Zoom: 1' 5' 1h 1d 5d 1m 3m 6m 1y Max
Num Tweets 292 | 10:55 March 01, 2011
All tweets
Feb 22
Feb 23
Feb 24
Feb 25
Feb 26
Feb 27
Feb 28
Mar 1
Zoom: 1' 5' 1h 1d 5d 1m 3m 6m 1y Max
Num Tweets 0 | 10:55 March 01, 2011
Tweets reporting an infrastructure damage

Problems in Statistical Network Analysis

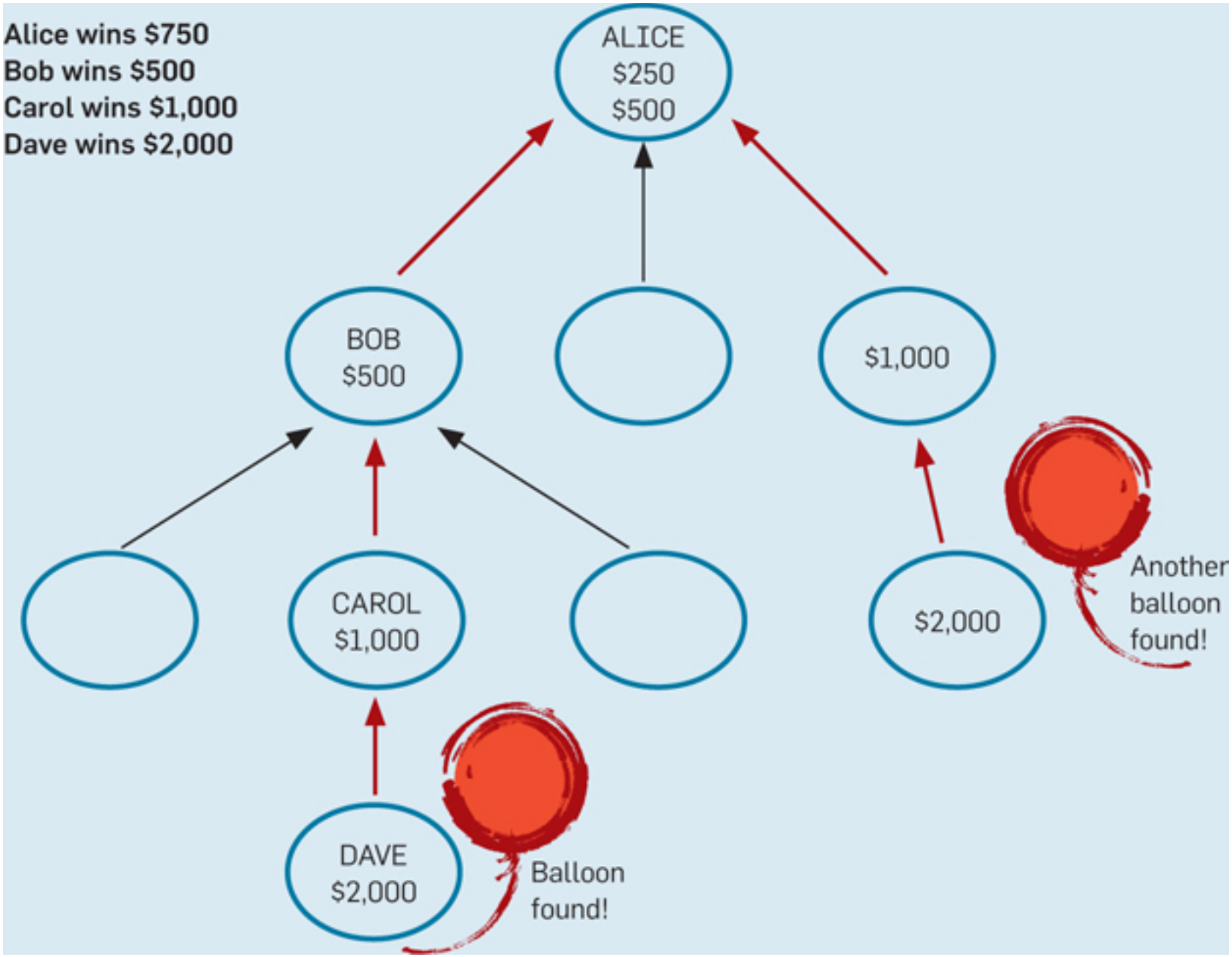
Dynamic Network Modeling

Incentive Design in Social Networks

- How to disseminate information in social networks?
- How to actively gather information?
- How to structure incentives?
- How to verify that the trust-worthiness of information?

DARPA Red Balloon Challenge

- Challenge: December 5, 2009
  - $40,000 to find 10 read weather balloons
- Result: found all 10 balloons with 9 hours


Alice wins $750
Bob wins $500
Carol wins $1,000
Dave wins $2,000
ALICE
$250
$500
BOB
$500
$1,000
CAROL
$1,000
$2,000
DAVE
$2,000
Balloon
found!
Another
balloon
found!

Red balloons: issues of autheticity

Typical real (left) and contrived (center and right) pictures of balloons.

http://cacm.acm.org/magazines/2011/4/106587-reflecting-on-the-darpa-red-balloon-challenge/fulltext

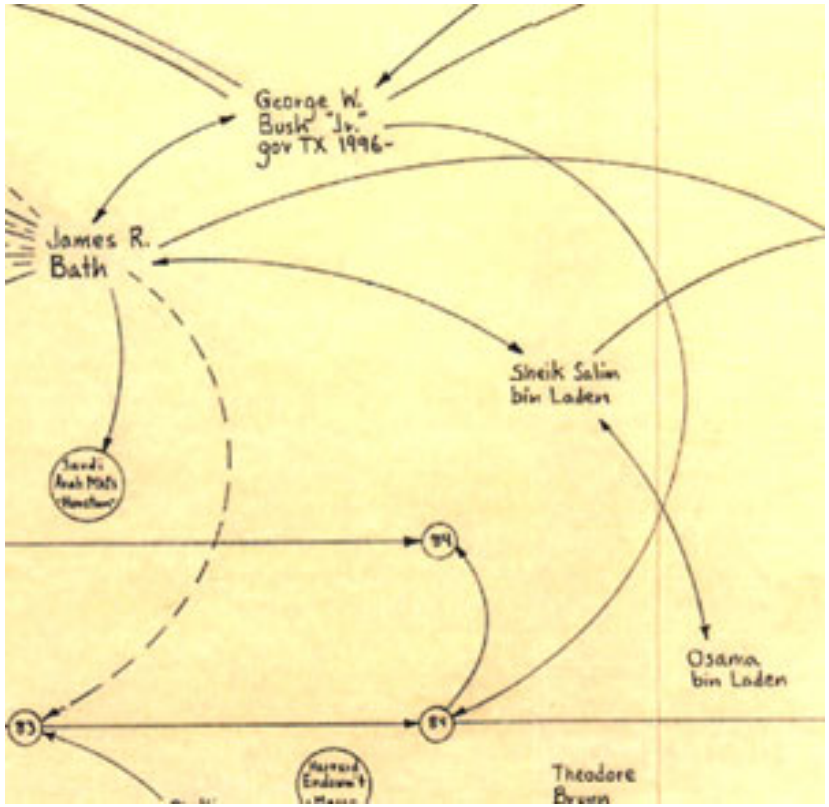
Network visualization

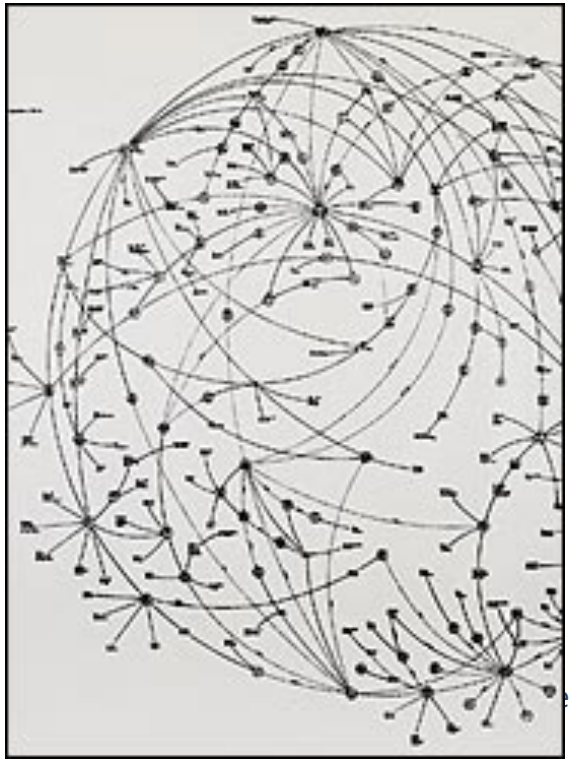
- How to make sense of a large network?
  1000s of nodes
  10,000s of nodes
  1,000,000 of nodes?

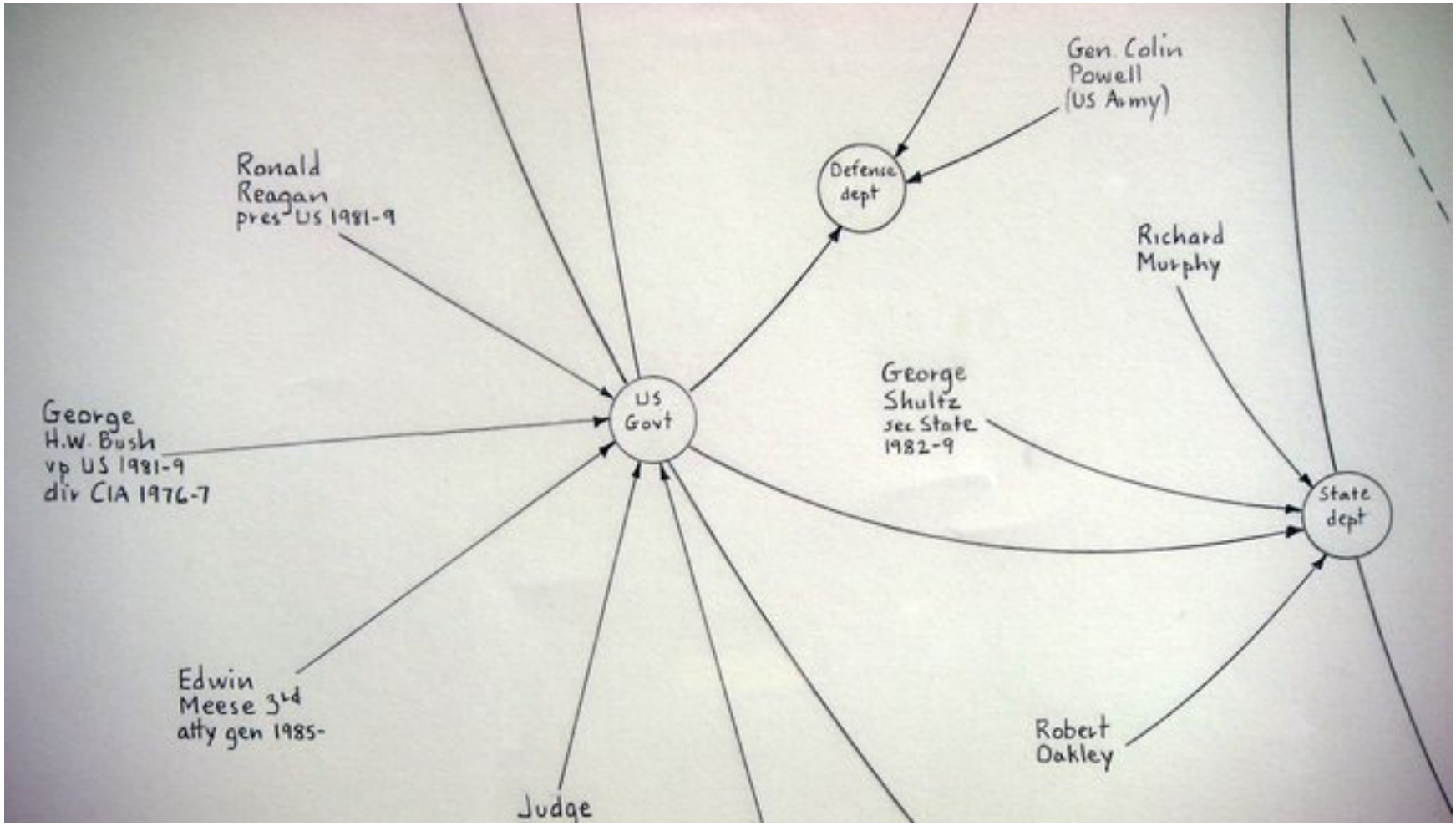
- What are efficient methods for displaying information about a network, manipulating it, and zooming in to reveal insight?

examples: Political/Financial Networks

- Mark Lombardi: tracked and mapped global financial fiascos in the 1980s and 1990s from public sources such as news articles







September 13, 2012 Mareike Wegener, NY Times
A detail of a 1999 Mark Lombardi diagram of the Iran-contra operation, from about 1984-86.

http://movies.nytimes.com/2012/09/13/movies/mark-lombardi-death-defying-acts-at-moma.html?_r=0

Understanding through visualization

- "I happened to be in the Drawing Center when the Lombardi show was being installed and several consultants to the Department of Homeland Security came in to take a look. They said they found the work revelatory, not because the financial and political connections he mapped were new to them, but because Lombardi showed them an elegant way to array disparate information and make sense of things, which they thought might be useful to their security efforts. I didn't know whether to find that response comforting or alarming, but I saw exactly what they meant."

  Michael Kimmelman

  Webs Connecting the Power Brokers, the Money and the World

  NY Times November 14, 2003

Slide by Lada Adamic, U Michigan

Summary

- What is CSS
- A collection of CSS problems
- This CSS class
  - Core Concepts (lectures)
  - Algorithms and Methods (lectures + labs)
  - Problems and Case Studies (lectures + projects)

- Coming up next: Representing and describing networks