Analysis of Social Media

MLD 10-802, LTI 11-772

William Cohen

9-11-12

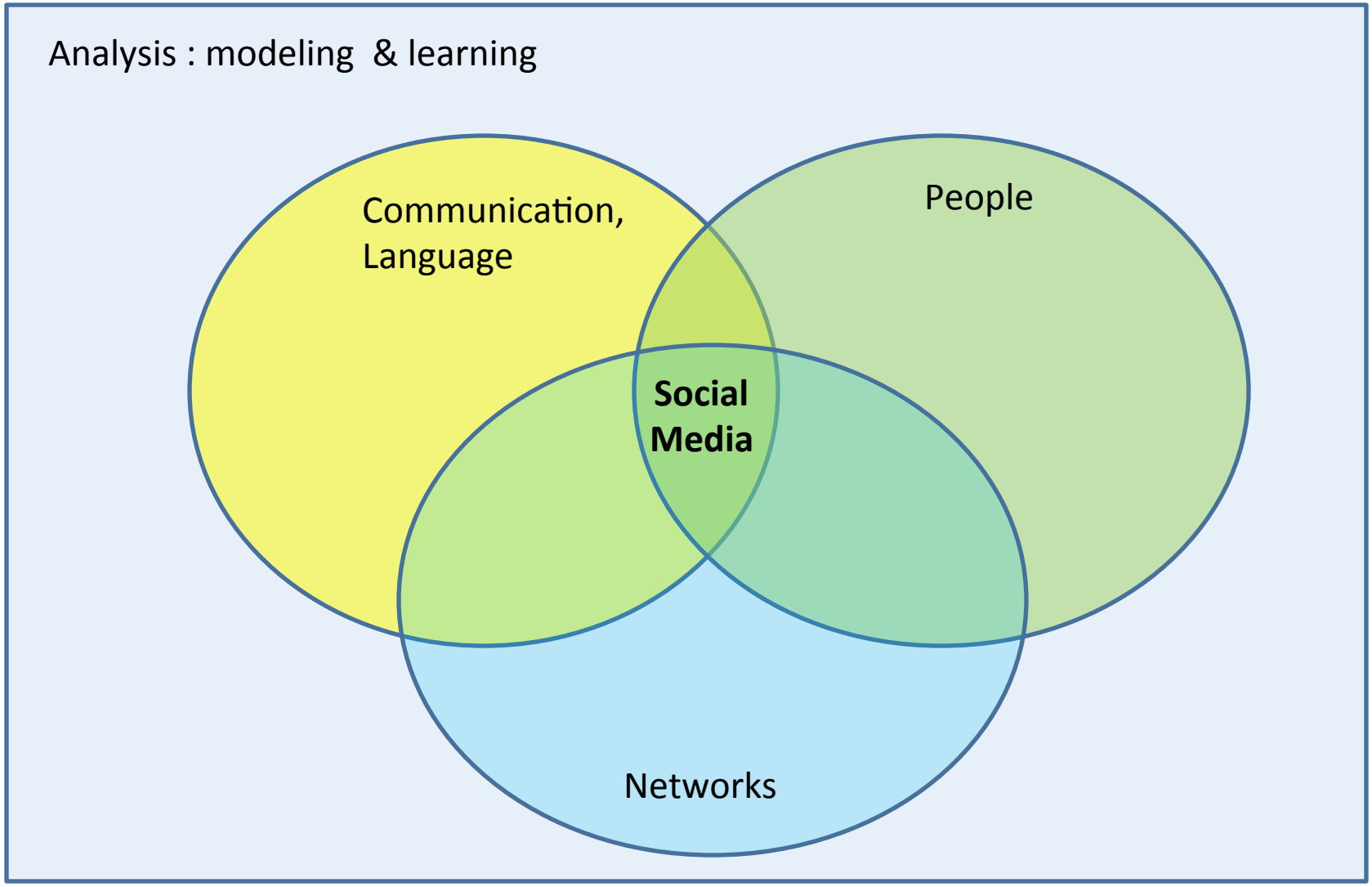
An example of social media

http://www.youtube.com/watch?feature=player_embedded&v=00KM53yZi2A

Outline

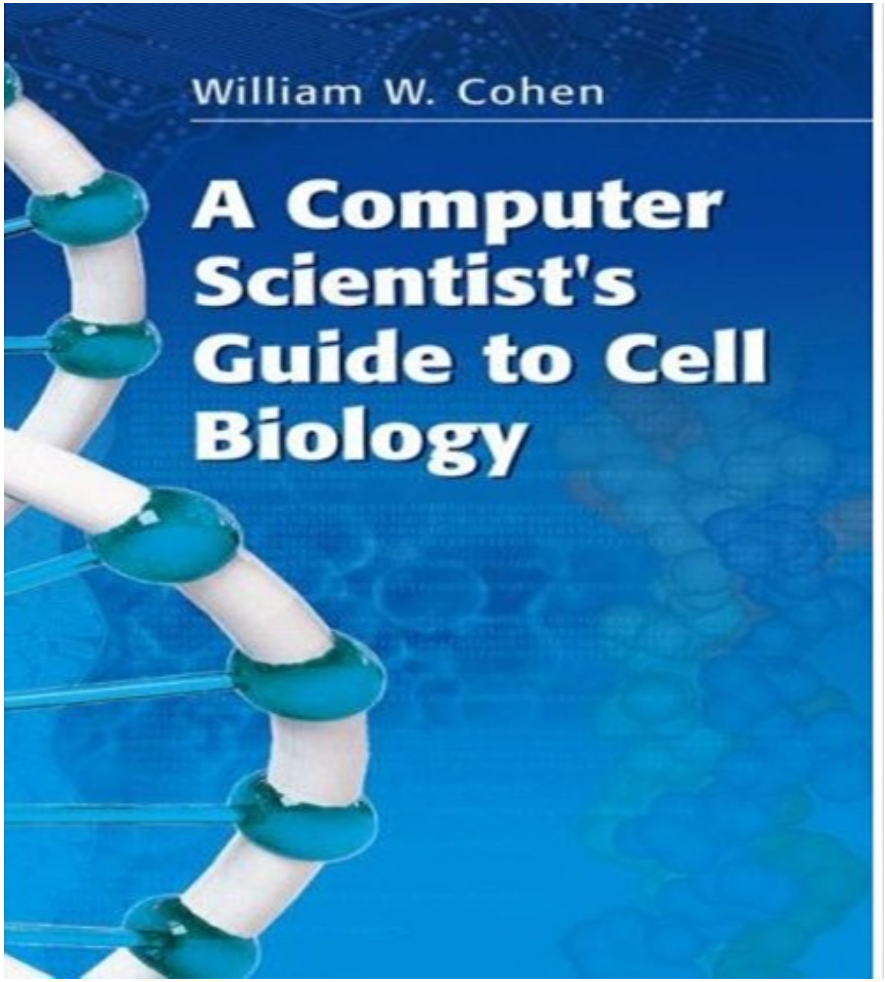
- What's the course about and why?
  - What's social media and why analyze it?
- Zoom out:
  - the landscape of "social computing" (context)
- Zoom back in:
  - where's the science? (topics we'll cover)
- Administrivia:
  - Preliminary syllabus
  - Projects and course wiki

What's this course about?

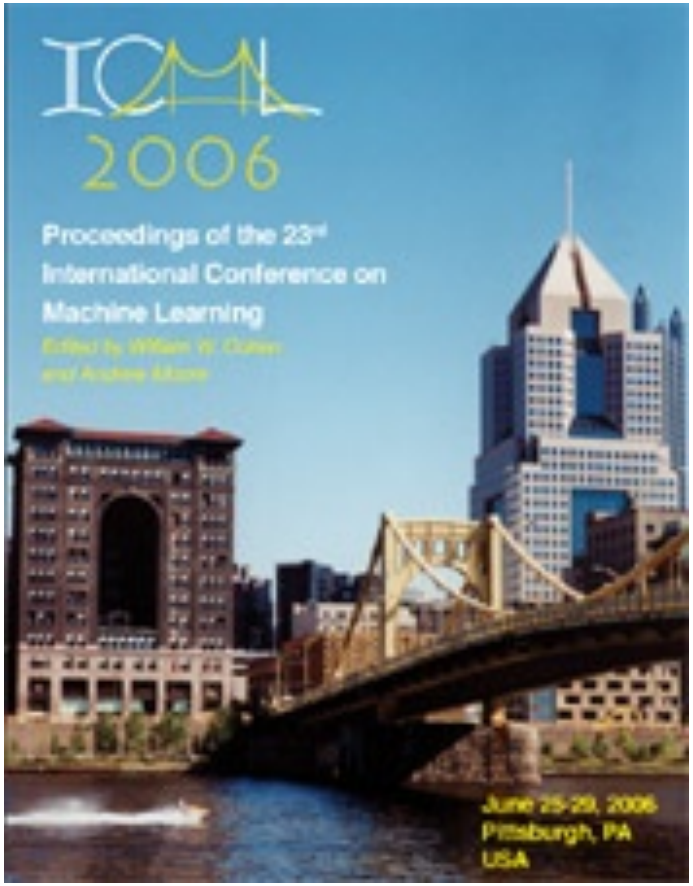


What is social media?

- Media **developed collaboratively** by a community
  - Examples:
    - Proc of the 23rd Intl Conf on Machine Learning
    - Wikipedia
    - YouTube
    - Blogosphere
    - WWW
  - Characteristics:
    - Decreasing cost and/or relative difficulty of participation → many more participants
    - Many participants → decentralized editorial process
    - Many participants → rapid changes over time

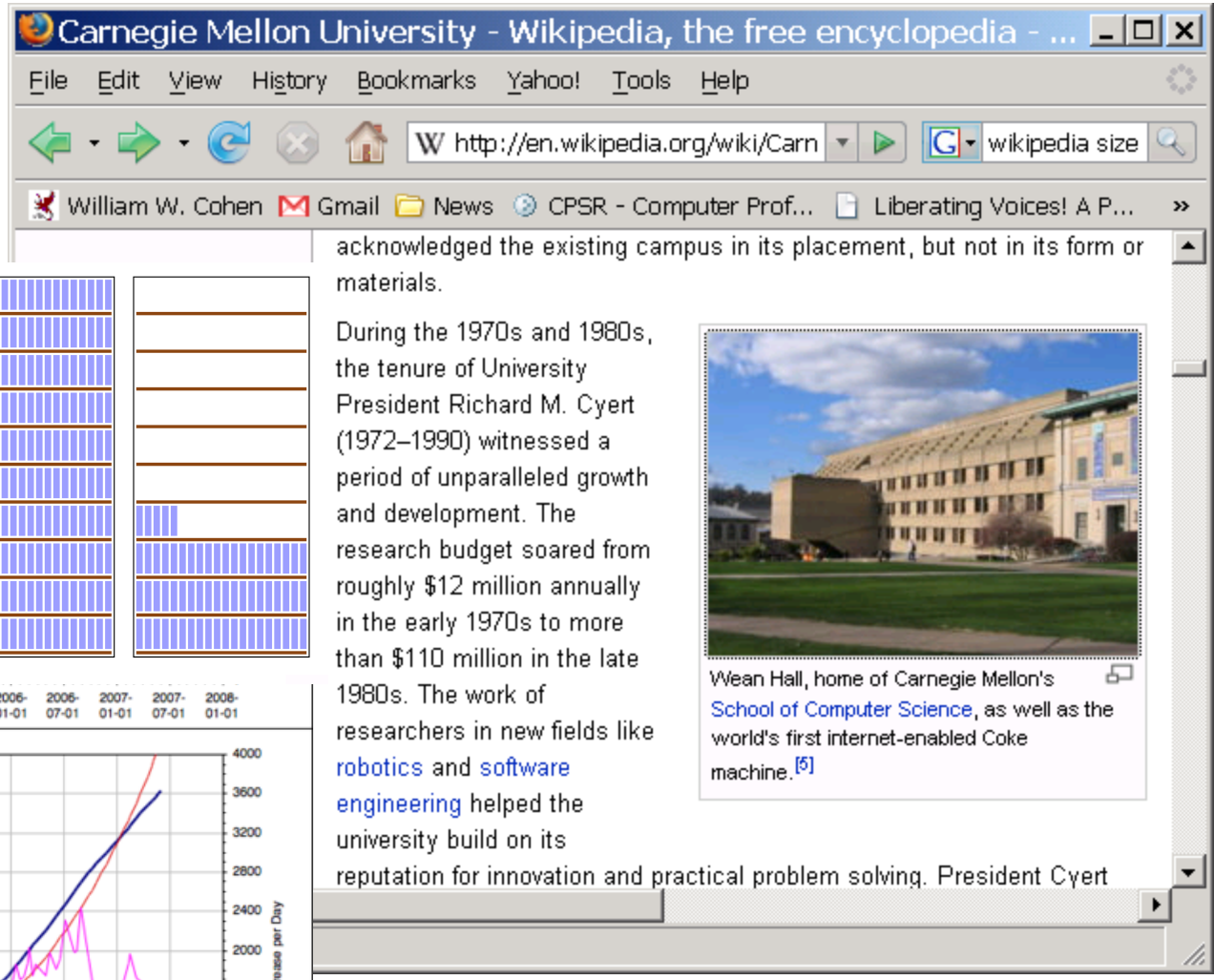


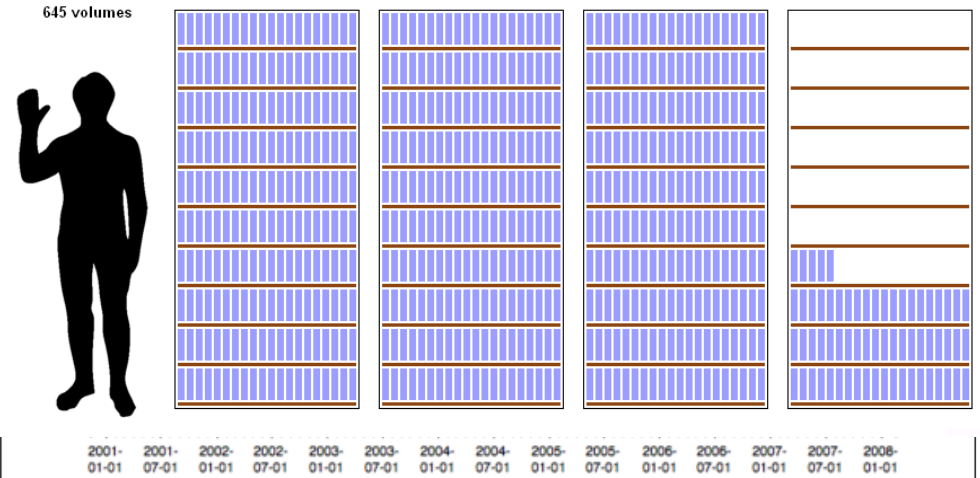
- O(1) editor
- O(1) reviewer
- O(1) author

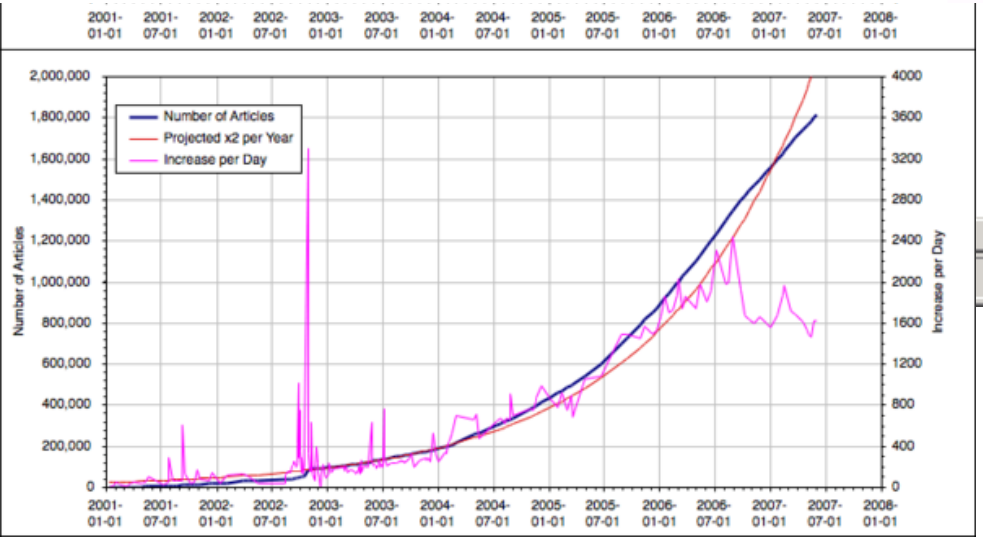


- O(1) editors (PC)
- O(10) SPCs
- O(100) reviewers
- O(1000) authors

c. 2007



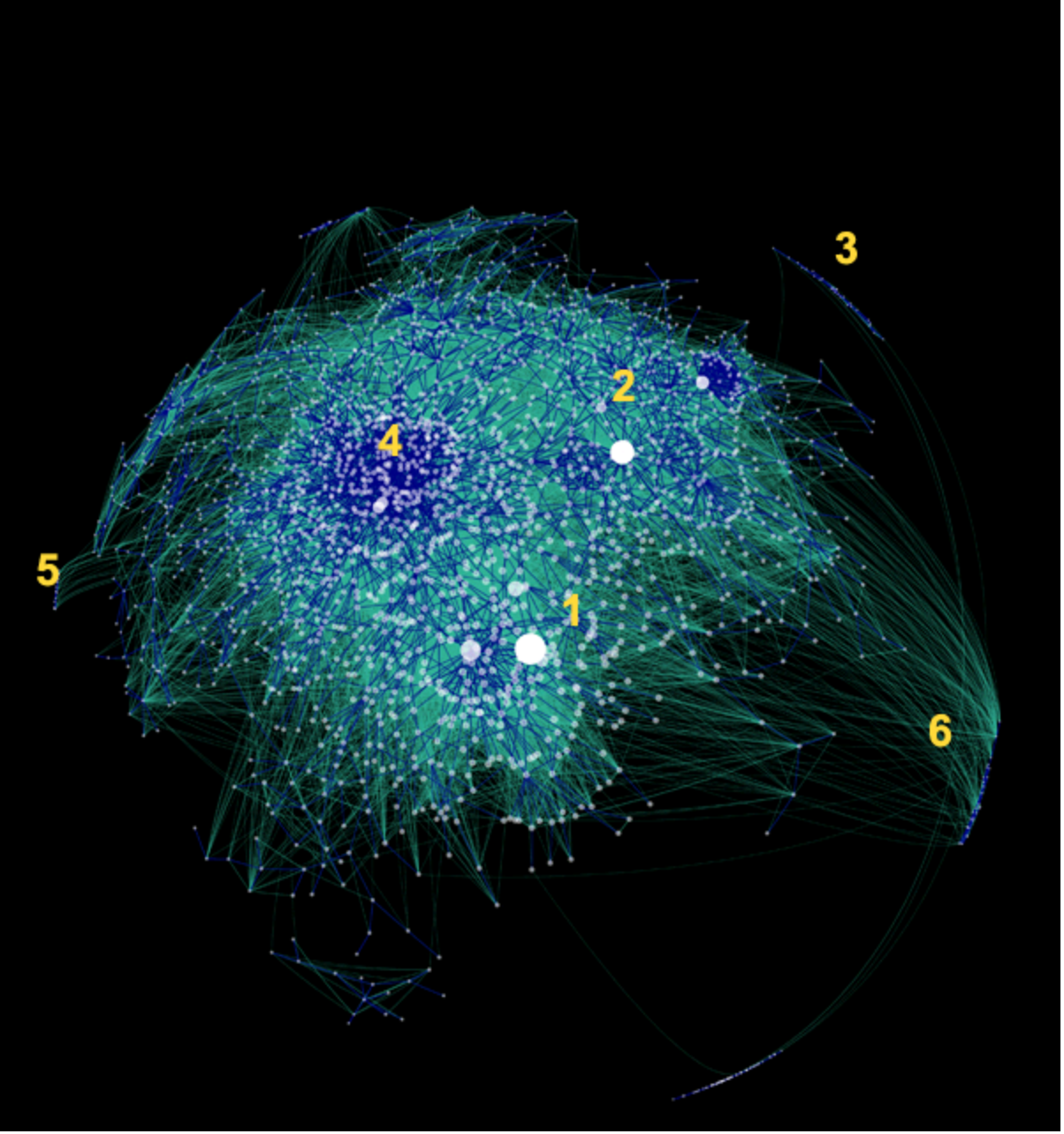




- many contributors
- editors & a consistent style

What is social media?

- Media **developed collaboratively** by a community
  - Examples:
    - Proc of the 23rd Intl Conf on Machine Learning
    - Wikipedia
    - YouTube
    - Blogosphere
    - WWW
  - Characteristics:
    - Decreasing cost and/or relative difficulty of participation → many more participants
    - Many participants → decentralized editorial process
    - Many participants → rapid changes over time



Visualization from Matt Hurst (Microsoft LiveLabs):

1. DailyKOS
2. BoingBoing
3. LiveJournal community
4. Reciprocally linked blogs (blue) around Michelle Malkin
5. Porn
6. Sports

NEWS
SEEK

Outline

- What's the course about and why?
  - What's social media and why analyze it?
- **Zoom out:**
  - **the landscape of "social computing" (context)**
- Zoom back in:
  - where's the science? (topics we'll cover)
- Administrivia:
  - Preliminary syllabus
  - Projects and course wiki

The bigger picture

- There are many kinds of *social technology* [M. Hearst]
  - Crowdsourcing
  - Idea Markets/Prediction Markets
  - Implicit Social Contributions
  - Shared Data
  - Shared World / Platform
  - Collaborative Creation
  - Social Networks

Crowdsourcing: Amazon's Mechanical Turk

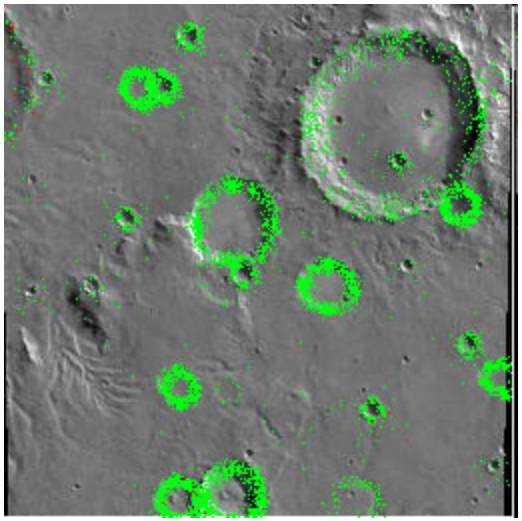
- A pool of thousands of people
- Small tasks, small pay
  - Many people do it for entertainment + pay
- Careful modularization required
- Widely used as a research tool
  - Relevance judgements for search
  - NLP assessments
  - User Interface assessments
  - …
- Eg, NAACL 2010 *Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*

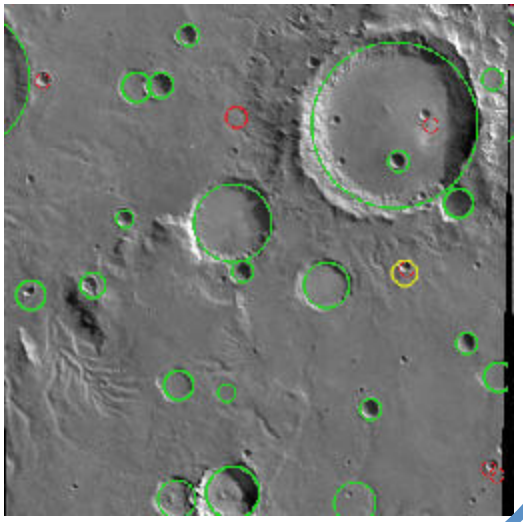
Crowdsourcing:
NASA Clickworkers

Early experiment, in 2001

Mars images from Viking Orbiter

Citizen Science in action





Internet Cat Video Film Fest draws 10,000 people

Posted by: Tom Horgen under Art, Movies | Updated: August 31, 2012 - 10:24 AM

10 comments | print

Recommend 211 Tweet 22 share +

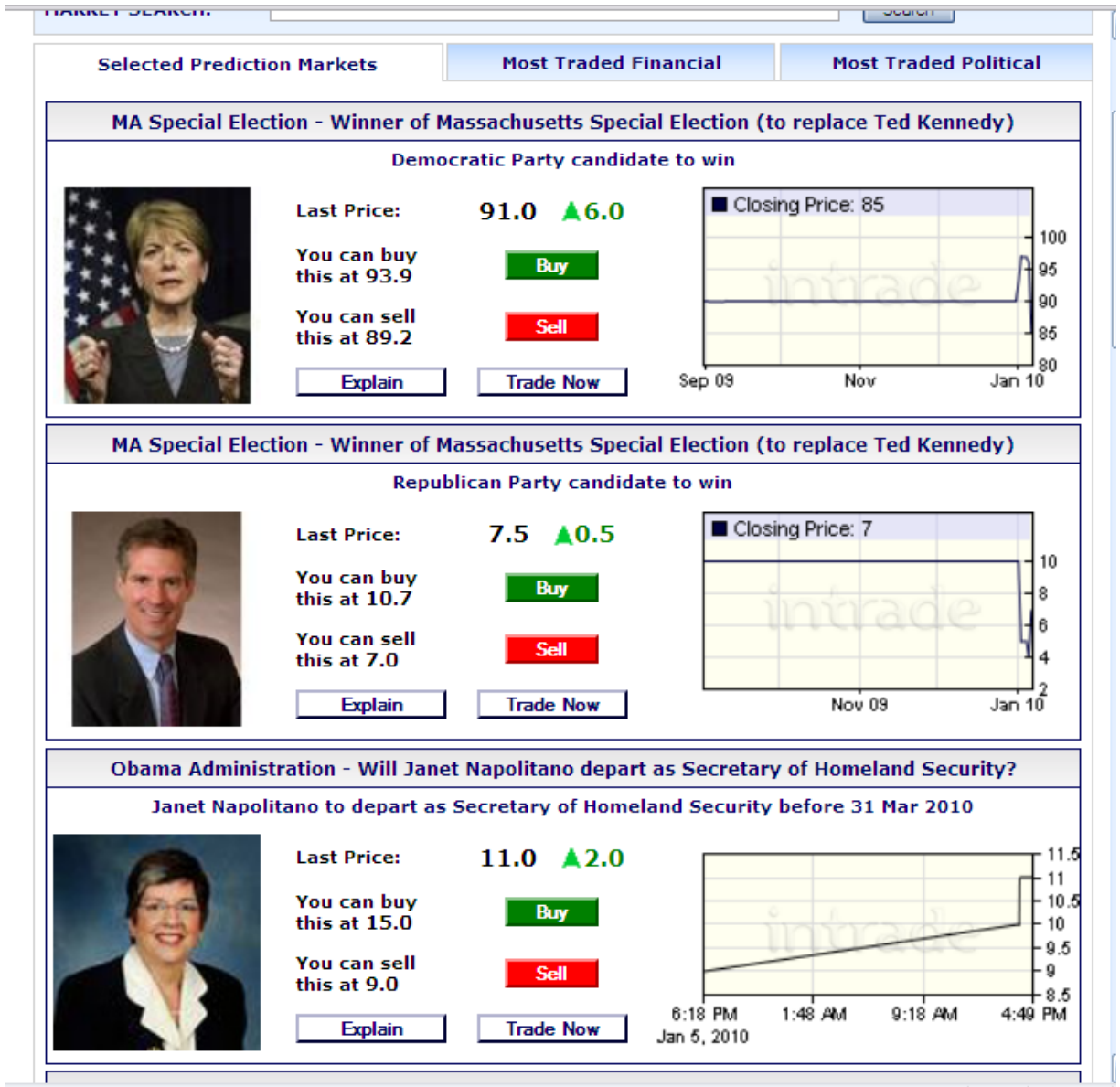
The numbers say it all: 10,000 people. That's how many cat-video lovers attended the **Walker Art Center**'s first **Internet Cat Video Film Festival** Thursday night. (And they weren't all cat ladies!)

That attendance figure rivals **Rock the Garden**, typically the Walker's biggest bash of the year.

Idea Markets and Prediction Markets

- Set up a market with an idea as a premise.
  - Public policy questions.
    - How would crime rates change if more citizens could legally carry hidden guns?
    - Make a market based on the crime rate change after a hidden-gun bill was passed. (Hansen 1999)
  - Internal product markets.
  - Manage IT portfolio via a trading market.

Idea Markets and Prediction Markets



Idea Markets and Prediction Markets: Key Points

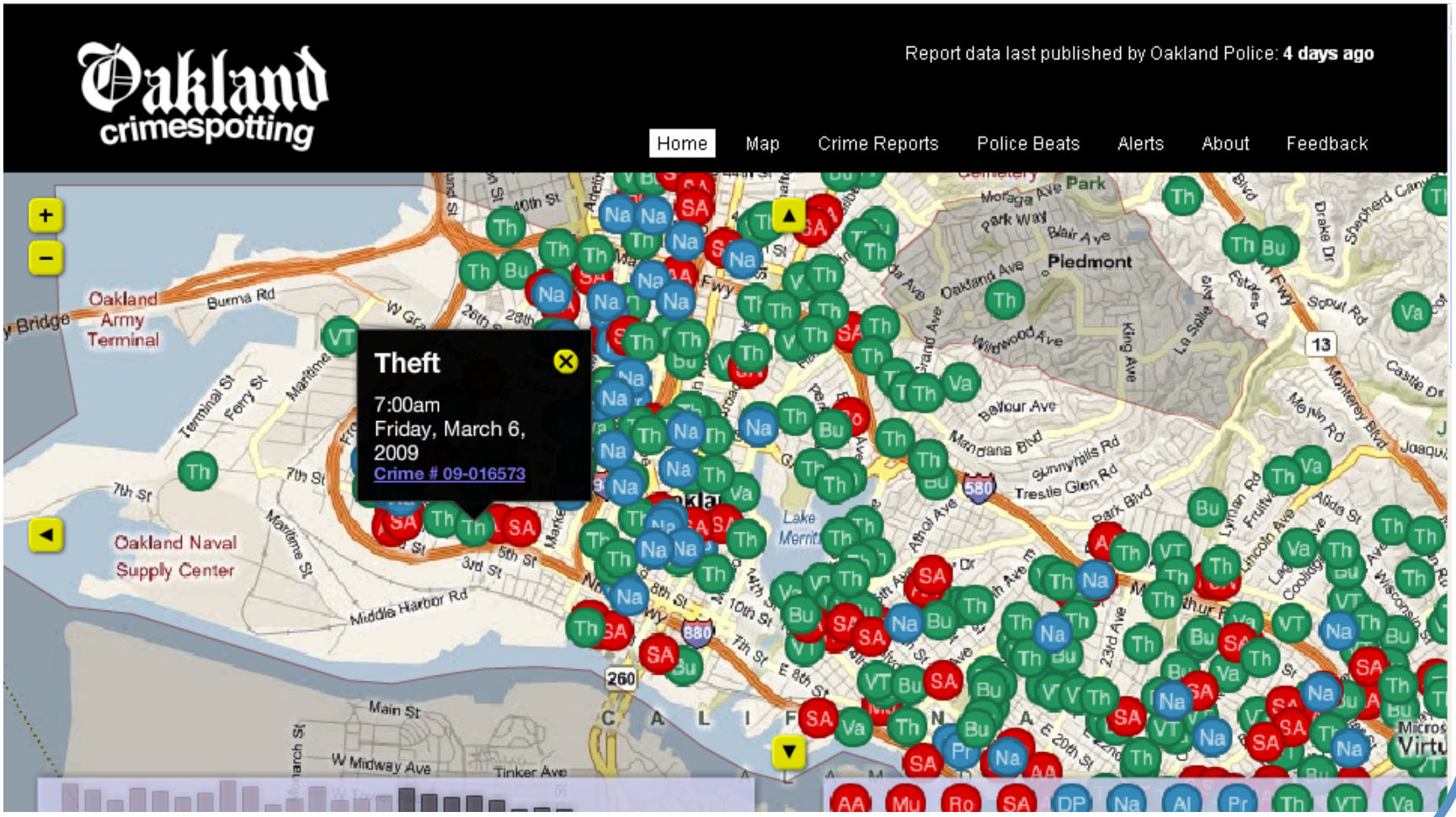
- Ferrets out hidden expertise or hidden information.
- People don't have to expose what they know directly.
- People don't have to know all pieces of the puzzle; it (hopefully) arises out of the mix.
- The connectivity of the Internet makes it possible like never before to find enough people with the right pieces of information to do this.

Implicit Contributions

- Clicks on
  - Search results
  - Recommended items
  - Ads
- Search queries
- Purchased items
- Anchor text (in hyperlinks)

Shared Data:
Augmenting Information Objects

Shared Data: Mashups



Shared Data:
Key Points

- Easy to participate, but may require some expertise or specialized access (have bought and used product, read the book, have an opinion about the legislation).
- Not a project with a coordinated goal; rather people are contributing to specific data items that they choose themselves.
- Being able to see and search the entire set of user-augmented data creates value for everyone.

Shared World / Platform:
Third Party iPhone and Facebook Apps

Large-Scale Collaborations

- Open source software
- Wikipedia
- Peer 2 Patent
- Science

PEERTOPATENT
in cooperation with USPTO

Sign Up | Login | Contact Us

COMMUNITY PATENT REVIEW

MY PROFILE | TUTORIALS | ABOUT P-TO-P

APPLICATION LIST | ARCHIVED APPLICATIONS | US PATENT CLASSIFICATIONS

Home > Applicatio

Active Patent Ap

Please click on th

Patent Application Ti

User-created metada interface

Platform for loyalty s

Market-based contin of use

Techniques for proje

PEERTOPATENT
in cooperation with USPTO

Sign Up | Login | Conta

COMMUNITY PATENT REVIEW

MY PROFILE | TUTORIALS | ABOUT P-TO

APPLICATION LIST | ARCHIVED APPLICATIONS | US PATENT CLASSIFICATIONS

Home

WELCOME TO PEER TO PATENT

Peer-to-Patent opens the patent examination proc
Become part of this historic program. Help the USPT
claims of pending patent applications. Become a co
patents.

- Click here to see a list of all applications.
- Click here to be notified of any new applications vi
- Click here to be notified about any new applicatior
- Learn more about how Peer-to-Patent works here

New Applications | Most Active Teams | Applic

	14	User-created metadata for managing i
	7	New! Risk assessment company
	5	Platform for loyalty services
	5	Smart, secured remote patient registra

more ...

PATENT APPLICATION PRIOR ARTIST AWARDS

These contributors submitted prior art or annotations used by the USPTO in making the determination of patentability.

Name	Type	Patent Application
Alexandre Eichenberger	Prior Art	Method and apparatus for an inductive doubling arch...
Gabriel Gomez	Prior Art	Image inversion
Sharat Mendu	Prior Art	Computer compliance system and method
Charles Peck	Prior Art	System and method for implementing a multi objectiv...
Jeff Morrill	Prior Art	Method of obtaining data samples from a data stream...
Christian Seifert	Prior Art	Honey monkey network exploration
Steven Pearson	Prior Art	Method and apparatus for selectively executing diff...
Susan Murray	Prior Art	Methods of enhancing media content narrative
Kathy Wang	Prior Art	Honey monkey network exploration
Mark Nowotarski	Prior Art	Tuning core voltages of processors
Walter Dietrich	Prior Art	System and method for retaining information in a da...

up to 400
lassificatio
718, 719, c
oftware an
ds and
ill be grant

TO REVIEW

urance
d casualty
ntitative ri
nce prem
a data m

Large-Scale Collaborations: Key Points

- Usually requires some expertise; the kinds of expertise needed are heterogeneous.
- People are working together towards a shared goal.
- Can only be done because of the supporting technology.
- The pieces need to be modularized (sometimes by a central entity).

Social Networks

- Undirected social networks
  - Facebook, MySpace, etc.
- "Directed" Social Networks
  - Connected within an organization, or for a purpose.
    - IBM's Dogear Intranet system
    - GovLoop
    - Slideshare

Directed social network: GovLoop

A message to all members of GovLoop - Social Network for Government

GovLoop now over 7,000 members. Super rad. Spread the word. Invite your friends.

Member of the Week: Barry Everett, EPA and Virtual Worlds guru
http://www.govloop.com/profiles/blogs/member-of-the-week-barry

Project of the Week: Mary Davie interviews Casey Coleman, GSA CIO, about her "Around the Corner" Blog.
http://www.govloop.com/profiles/blogs/project-of-the-week-around

It's Out! GovLoop presents the I Am Public Service e-book.
http://www.slideshare.net/iampublicservice/i-am-public-service-1102253?src=embed

March Madness Starts in a week! Join the GovLoop Fantasy Challenge. Password to join is "govloop"
http://tournament.fantasysports.yahoo.com/t1/register/joinprivategroup assign team?GID=23430&P=govloop

1st GovLoop podcast is out. GovLoop picks the brain of Andy K about I Am Public Service:
Stream on GovLoop.com home or download from iTunes - itpc://recordings.talkshoe.com/rss41639.xml

Great new blogs on topics including Chris Anderson on Gov 2.0, Facebook is for Children, Congress Delivers Social Media Failure, and Suing Social Media into Silence http://www.govloop.com/profiles/blog/list?pageSize=10

GovLoop

Can Government Procurement Be Streamlined By Using Collaboration Technologies and Social Media?

By Dennis D. McDonald, Ph.D. *Author's note: this is a republication from the author's web site located here*. The report Six Practical Steps to Improve Contracting by Dr. Allan V. Burman, Adjunct Professor, George Mason University, is based on a series of discussions co-sponsored by The IBM Cente… Continue

Added by Dennis McDonald on February 27, 2009 at 3:22pm — 5 Comments

In which I make a case for (a little bit of) Web 1.0 in the Government 2.0 world

I've been thinking about Dennis McDonald's thoughts about K-TOC. He wrote: "I guess I see an advantage to being able to easily differentiate between a web site that serves as an official portal, and a web service that facilitates a mix of formal and informal communication. The question is, how realistic is it to combine the two?" I tap-danced around Dennis's points in my response, but I've had a chance to think about it for… Continue

Added by Patrick Quinn on February 25, 2009 at 4:34pm — 3 Comments

Making Recovery.gov first step toward smart regulation

(This post appeared originally in the Huffington Post, Feb. 23, 2009. Reprinted with permission)

The Obama Administration created Recovery.gov as a critical stimulus component, taking a "don't trust us, track us" approach to assure funds are

June (21)
May (5)
April (4)
February (1)
2007
November (2)

GovLoop

Last View: 3:15PM, 13 Mar 09
Last On: 7:11PM, 13 Mar 09
Mary Davie
Female
Arlington, VA
United States
Add as Friend

MY RANKING

Super User
1320 points

Share
Block Messages

Applications
Blog Posts (1)
Discussions (5)
Events
Groups (7)
Photos
Photo Albums
Videos

MARY DAVIE'S FRIENDS

Bjorn Miller left a comment for Mary Davie yesterday

Lesa Scott commented on the blog post 'Project of the Week - "Around the Corner"' yesterday

"Mary, now you've done it, you've got me thinking about becoming a blogger!"

Martha Przysucha commented on the blog post 'Project of the Week - "Around the Corner"' on Thursday

"Thank you Mary, and Casey! This is an excellent intro to those considering launching a blog or just trying to understand social media. I'm interested in the measurements for ROI on social media implementations. Will watch for more from GSA!"

Janice commented on the blog post 'Project of the Week - "Around the Corner"' on Wednesday

"Very informative Mary. I'm learning a lot about blogging and social networking thanks to you. I'm having fun too."

Greg Berry commented on the blog post 'Project of the Week - "Around the Corner"' on Wednesday

"Great job Mary! Thank you for "introducing" me to Casey. The timing couldn't be better, we are launching our new blog soon (since Google deleted our old one for reasons unknown to us) and it gave me some ideas on how to structure it..."

Mary Davie and Faye Farah are now friends on Wednesday

Adriel Hampton commented on the blog post 'Project of the Week - "Around the Corner"' on Wednesday

"Thanks, Mary! Always interested in the process behind official social media projects."

GovLoop

Marti Hearst Sign Out Search Social Network

govloop

connecting the government community

Main Invite My Page Members Rankings Blogs Groups Forum Jobs Events Multimedia Knowledge

All Groups My Groups

+ Add a Group

All Groups (288)

Search Groups

Sort by: Most Active

Geeks in Government
120 members
Latest Activity: 4 hours ago
Science and sci-fi geeks come out of the government closet!

CRM
4 members
Latest Activity: 1 day ago
govloop group dedicated to educating on the topic of CRM (Customer Relationship Management). Collaborative group offering best practices, findings, li...

Mac Lovers Unite
31 members
Latest Activity: 1 day ago
To all mac lovers - long-time fans and recent

Web Metrics
84 members
Latest Activity: 1 day ago
Discussion group for Government Web

Marti Hearst
Sign Out
Inbox
Alerts
Friends – Invite
Settings
Quick Add...

Social Networks: Key Points

- Usually no expertise required.
- Identity is central, relationships are key.
- People make contributions, or "just hang out."
- Value rises out of *connectedness*, sometimes leading to virality.

The bigger picture

- There are many other kinds of *social technology* [M. Hearst]
  - Crowdsourcing
  - Idea Markets/Prediction Markets
  - Implicit Social Contributions
  - Shared Data
  - Shared World / Platform
  - Collaborative Creation
  - Social Networks
    - "Directed" and "undirected"

Parallel , independent processing

Shared prices data, APIs, …

Shared creative goals, created artifact

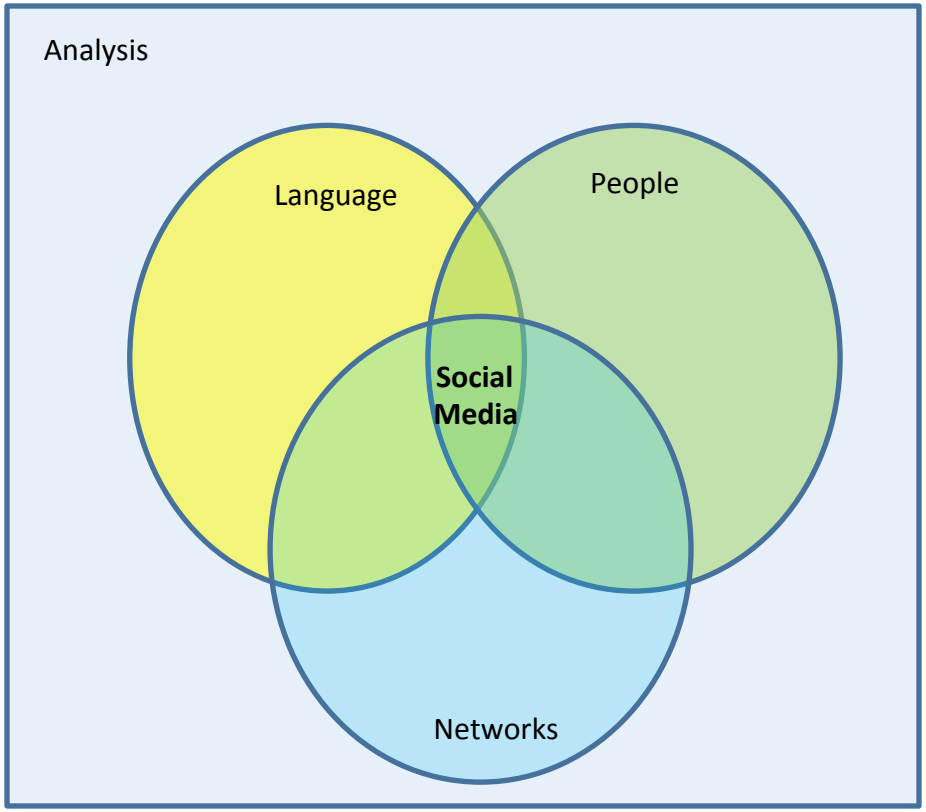
Linguistic communication

Outline

- What's the course about and why?
  - What's social media and why analyze it?
- Zoom out:
  - the landscape of "social computing" (context)
- **Zoom back in:**
  - **where's the science? (topics we'll cover)**
- Administrivia:
  - Preliminary syllabus
  - Projects and course wiki

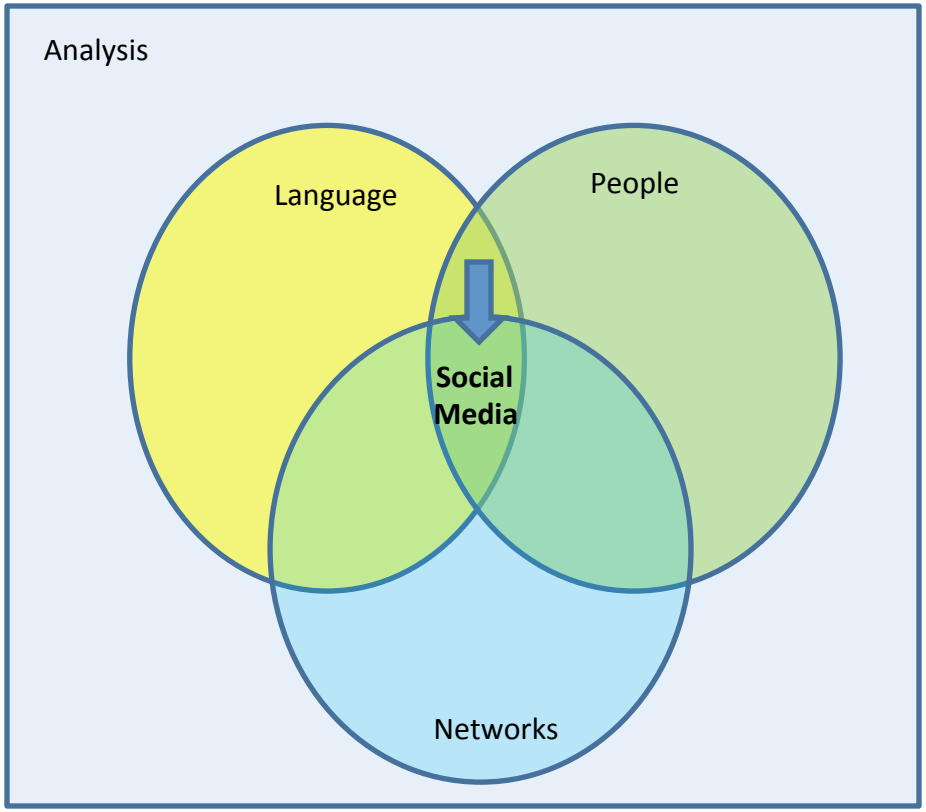
What's this course about?

- This is an *emerging area*, not a mature one.
- The problems and techniques are not well-understood.
- A lot of what we will cover is from areas on the *edges* of this picture, working in…



Research questions & areas

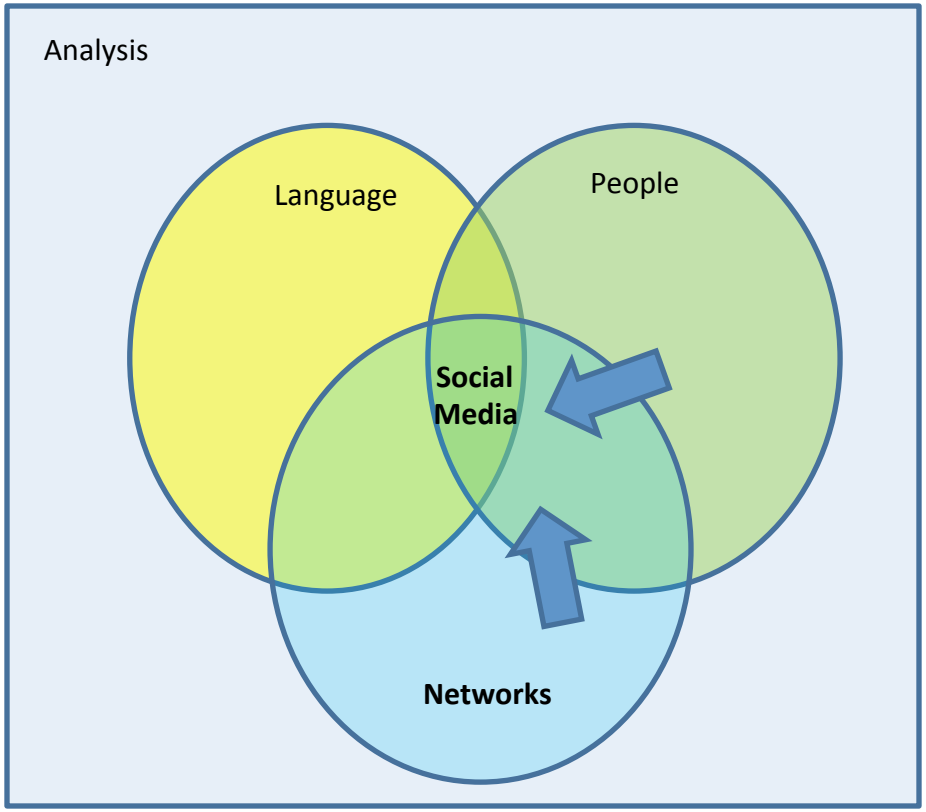
- How and why do people communicate in social settings?
- ➔ Understanding the language of *sentiment* and *opinion*.
- ➔ How language affects *behavior* ("shallow pragmatics") in social settings.



[e.g., email and message-board response behavior; commenting activity on blogs]

Research questions & areas

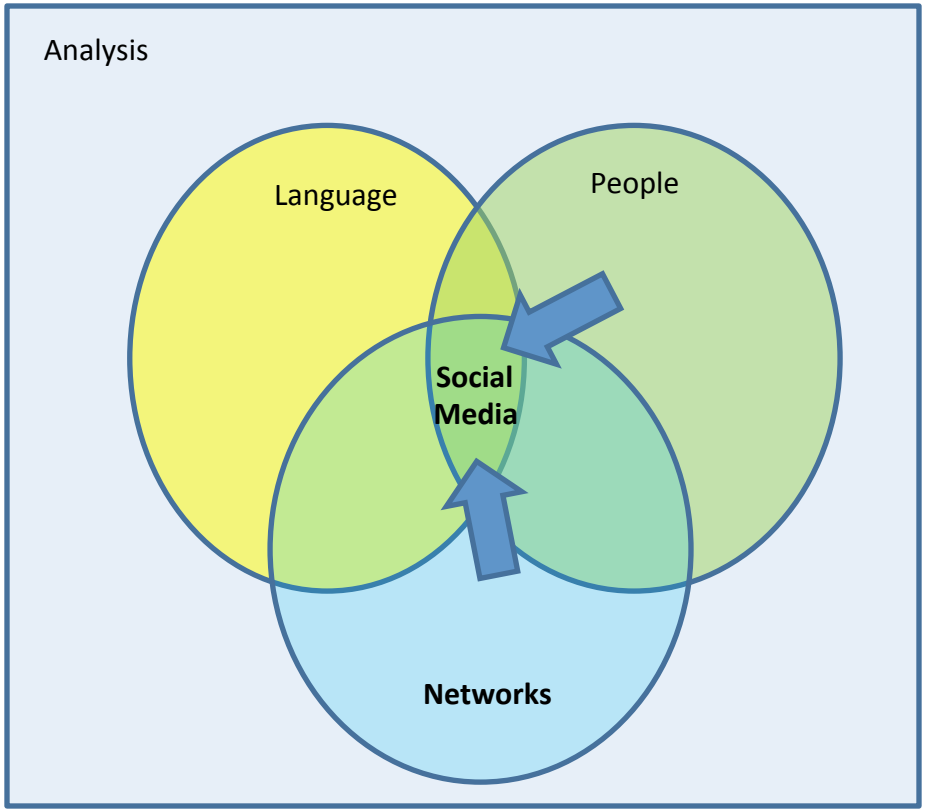
- What sort of social networks do people form? When do they choose to communicate? When are they effective?
- ➔ Social network analysis; homophily; small-world phenomena.
- ➔ Probabilistic models for structure in graphs.
- ➔ Other models for structure in graphs (e.g., spectral, modularity, …)



["Networks Crowds & Markets", Ealey & Kleinberg – mix of economics, psychology, graph theory; "Networked Life", Kearns

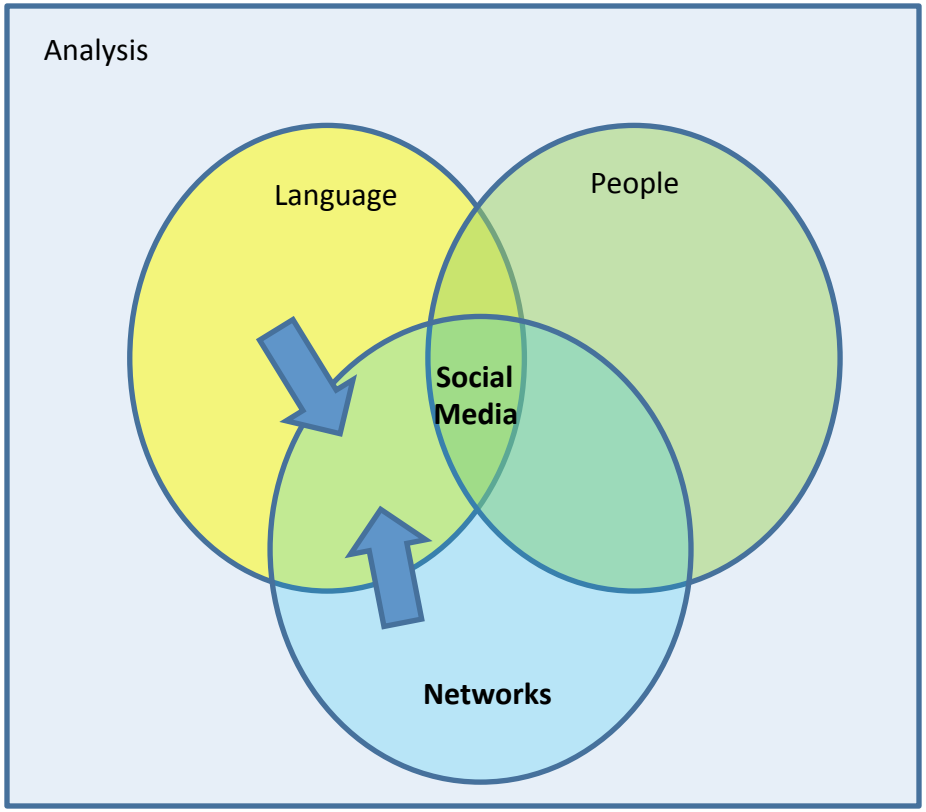
Research questions & areas

- What sort of social networks do people form? When do they choose to communicate? When are they effective?
- ➔ Collaborative annotations (folksonomies) and collaborative rating schemes.



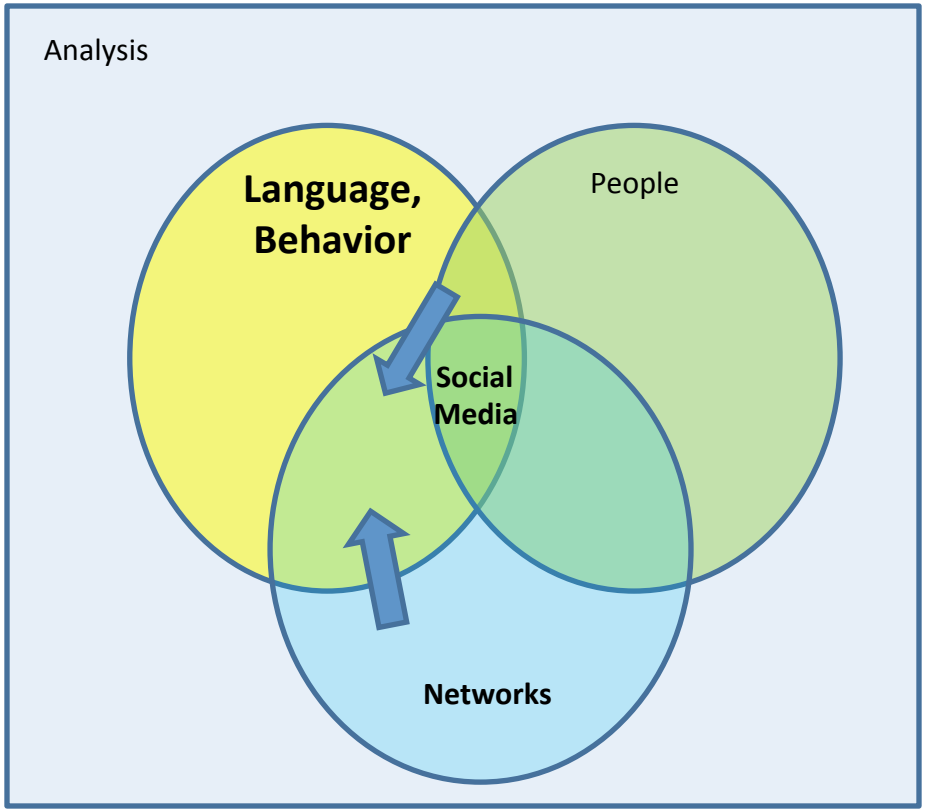
Research questions & areas

- How do you model collections of documents in graphs? How do ideas spread through a network?
- ➔ Hybrid models of text and connections [Relational LDA]
- ➔ Models of diffusion and influence.



Research questions & areas

- How do you model collections of documents in graphs? How do ideas spread through a network?
- ➔ Hybrid models of text and connections [Relational LDA]
- ➔ Models of diffusion and influence.
  - Viral marketing
  - Collaborative problem-solving.



What isn’t “analysis of social media?”

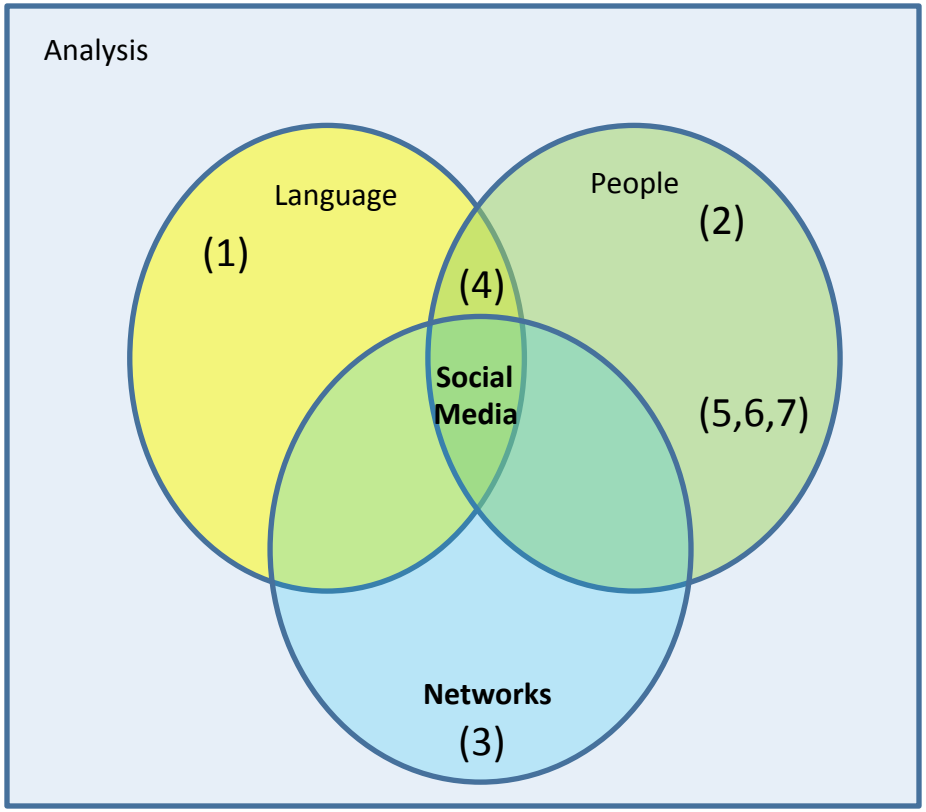
What isn't "analysis of social media?"

1. Animal signaling (nonhuman proto-languages).
2. Behavorial psychology of individuals.
3. Network models of protein-protein interaction

4. Event extraction from AP news.
5. Crowdsourcing (AMT)
6. Games with a purpose.
7. Prediction markets.

for NLP

Of course ideas and techniques from any of these areas might be relevant….

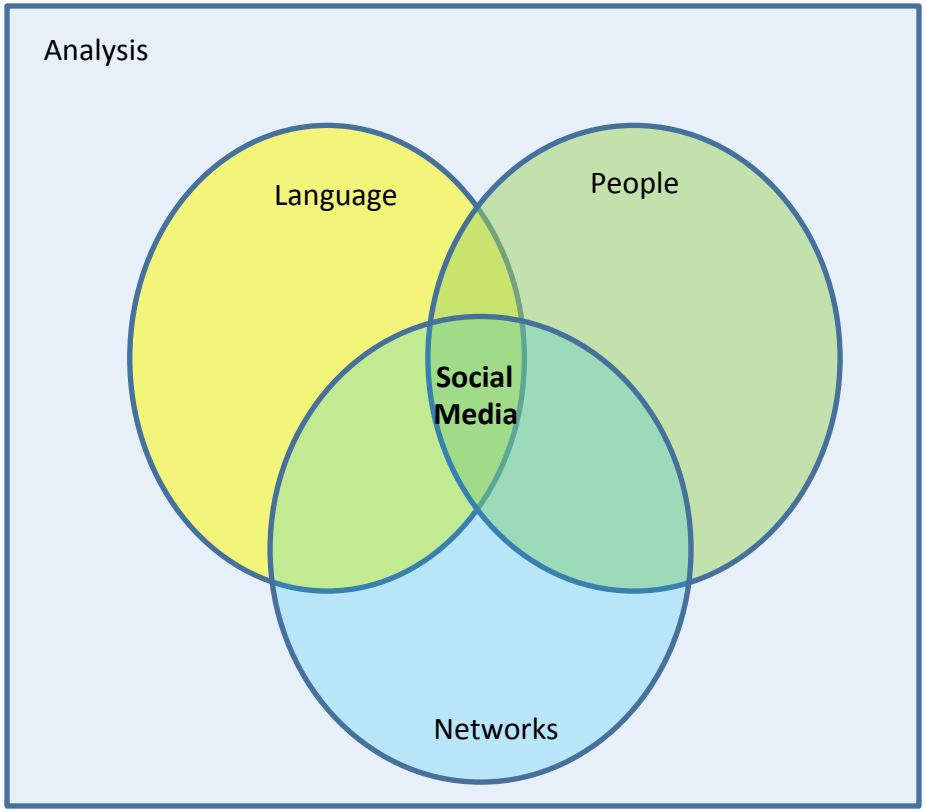


Outline

- What's the course about and why?
  - What's social media and why analyze it?
- Zoom out:
  - the landscape of "social computing" (context)
- Zoom back in:
  - where's the science? (topics we'll cover)
- **Administrivia:**
  - **Preliminary syllabus**
  - **Projects and course wiki**

Syllabus

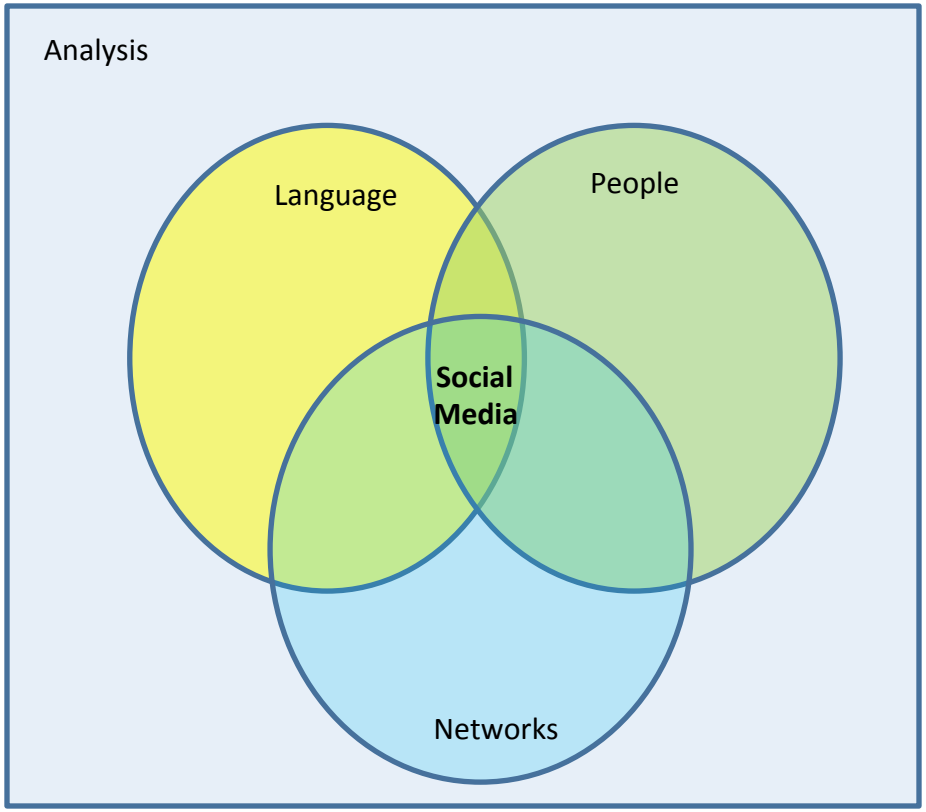
- 1. Background(6-8 wks, mostly me):
  - Opinion mining and sentiment analysis. Pang & Li, FnTIR 2008.
  - Properties of social networks. Easley & Kleinberg, ch 1-5, 13-14; plus some other stuff.
  - Stochastic graph models. Goldenberg et al, 2009.
  - Models for graphs and text. [Recent papers]



Syllabus

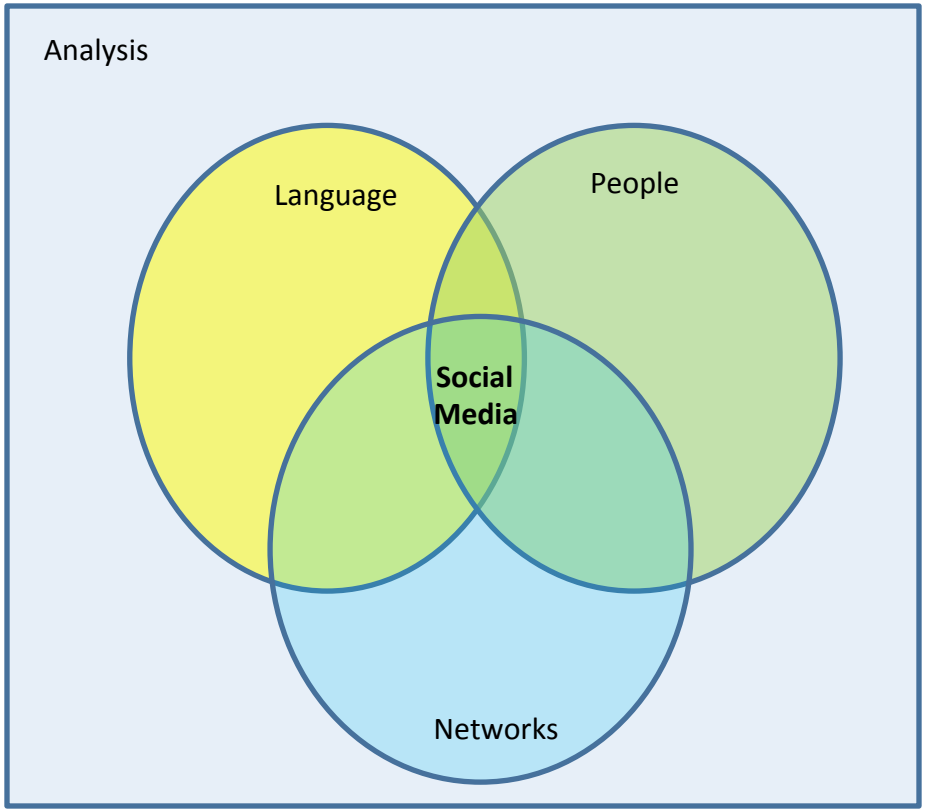
- 2. Special topics (mostly me)
- 3. Review of current literature (mostly you, via reviewsß).

- Asynchronously: guest talks from researchers & industry on sentiment, behavior, etc.



Syllabus

- Guest talks:
  - Cosma Shalizi (stats): Causation and correlation in networks
  - Jan Weibe (Pitt): sentiment analysis
  - Justin Cranshaw (HCII): LiveHoods
  - Jen Mankoff (HCII): StepGreen
  - Ramnath Balasubramanyan (LTI): LDA-like network models
  - Yi-Chia Wang (LTI/HCII): Engagement in Online Communities



Class project and assignments: background

- Spring 2007: We all built a wiki summarizing the field as of then:
  - http://www-2.cs.cmu.edu/~wcohen/10-802/fixed/Main_Page.html
  - Each student contributed K things to the wiki (mostly paper writeups) as well as presenting.
  - No projects (6 credits)
- It was a nice communal goal for the class to have.
- It was a great resource
  - but wasn't backed up properly and a lot of it was lost ☹
- Fall 2009: in IE course, students did writeups and posted them on wikis
  - But the bar for a writeup was way lower.
- Spring 2010, 2011…..

Class project and assignments: motivation

- My secret plan:
  - Comprehensive graph relating methods, problems, datasets, and papers.
  - Each page is an *edge* (or planning operator) from something(s) you know to something(s) you don't.
- Uses:
  - Personalized summarization of a new paper based on what you know: "This is a nonparametric variant of the topics of time model applied to the NIPS papers dataset and a subset of the ICWSM 2007 blog data, used to solve the task of retrospective event detection."
  - Collective summarization of a subfield …
  - …

Presentations and Writeups: 2010 and 2011

- Generally a paper discusses a problem and one or more methods applied to that problem.
  - Problem: dataset and evaluation criteria.
  - Method: some algorithm.
- Part 1:
  - Post the appropriate number (4-6) of wiki pages discussing the methods, problems, & datasets used in the paper.
  - …plus a summary of the paper in terms of these.
- Part 2:
  - Give a 20-30min talk presenting the work to the class…hopefully when we're discussing related work in lectures.

The project: 2010 and 2011

- Project is defined by a problem and one or more methods.
  - Problem: dataset and evaluation criteria.
  - Method: some algorithm.
- Phase one:
  - Propose your (joint?) project to me: Feb 1
- Phase two:
  - Post the appropriate number (8-10) of wiki pages for the work that is most related (methods , problems, or data).
  - Get the data and analyze it.
  - Give a talk on this (after spring break).
- Phase three:
  - Do the project and write it up (like a conference paper).

We're doing things differently this year…..

- There are too many of us to all present
  - So: no presentations
  - But: I do still want us to interact….
- Old assignments are limited
  - My "structured wiki" doesn't capture nearly all aspects of research
  - What else is there besides methods, datasets, papers, problems?
- The project plan:
  - We'll explore and discuss some options together
  - Constraints:
    - incremental progress toward your project
    - get feedback on your understanding of what you're doing early
    - *create something of future value* - not busywork

Grading

- Class participation (10%):
  - 20%: read the material in advance, come with questions/ comments in mind; contribute to the class wiki; …
- Project preparation (40%):
  - this will be a series of wiki-based, paper-review sort of homeworks
  - we will have speed-dating project presentations the week of 11/12, where *you describe your data and a baseline system*
- Project (50%):
  - 15% talk
  - 15% format and presentation
  - 20% quality of the research and results

Assignment 1

- Due: next Tuesday 9/20
  - Get an account on the wiki (krivard@cs, unless you already have one)
  - Create a user home page with
    - your name
    - your picture (optional)
    - what you want to get out of the class
    - a sentence or two about what you'd like to do for your project (non-binding, just for my interest)
- Due: this Sun 9/18
  - Give me feedback on the proposed format for assignments, using the wiki