LECTURE 2:SOCIAL MEDIA AND NETWORKS

1

COMP4641: Social Information Network Analysis and Engineering
Friday February 6th 2015

Social media

2

- Social media has transformed society
  - Reduced barriers to communication
  - Democratized content publication
- As a computer scientist…
  - Tend to ignore users
  - Social media makes users a part of the system
- Important to understand interactions
  - Within the system (traditional CS)
  - Between users and system (HCI)
  - Among users themselves (sociology)

What is social media?

3

- Systems with user interaction as critical component
- Online communities
  - Facebook, MySpace, YouTube
- Communication systems
  - Skype, Instant Messaging
- Social news media
  - Blogs, iReport
- Online worlds
  - World of WarCraft, Second Life

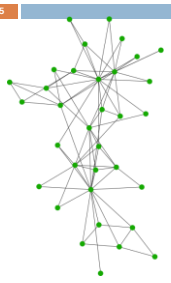
Why is social media interesting?

4

- Two reasons (to me):
- 1. Observe social interaction at scale
  - Social media based user interactions
  - Scale not possible before
- 2. Relate information and people
  - Online social networks now content-sharing systems
  - Can attach reputation of users to content

1. Observe social interactions

5



Anyone recognize this network?
- Zachary's Karate Club

- Collecting it involved massive field work
  - Manually observe people
  - Trace interactions for two years (!)
  - Will discuss more later
- Limit in scalability of this approach
  - Biases from interviewing
  - Time spent

Opportunity: Large-scale data

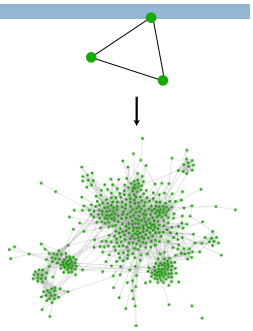
6

- An opportunity to scale up observations
  - "Field work" required may be reduced
- Social media sites have complete history record
  - Interactions, discussions, friendship creation (and deletion), …
  - Entire evolution of a group of users
- At incredible detail
  - 76% of Facebook users login at least once a day
  - 20B people-minutes spent on Facebook per day
  - Every interaction recorded

The curse of scale

7

- Scale is both a blessing and a curse
- Blessing
  - Confidence in results
  - Certain effects only seen at scale
- Curse
  - Miss many local interactions
  - Links "mean" less
  - Comparing networks hard
- Important to keep limitations in mind



2. Relate information to people

8

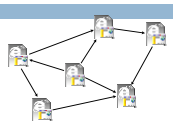
- Popular way to connect and share content
  - Photos, videos, blogs, profiles, news, status...
  - Twitter (500 M), Facebook (1.393 B)
- Growing exponentially
- Incredible amounts of content being shared
  - Facebook (350 M photos/day)
  - YouTube (100 hours of video/min)

myspace.com a place for friends · facebook · YouTube Broadcast Yourself · orkut · flickr · twitter

A new way of organizing information

9

- Web organized with content-content links
  - Link structure exploited (e.g., PageRank)
- Social media organized using
  - User-user links (social network)
  - User-content links (favorites, etc)
- New platform for information sharing



Web

Social networks

Relates information to people

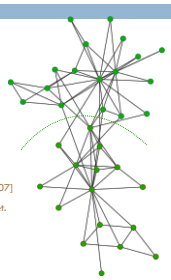
10

- Today, social network used to structure information
- Can we extract other information?
  - Combination of who and what very powerful
- Social network connects content with
  - (Multiple) user's reputation
  - Community the user is part of

But why study networks?

11

- Does network science make sense for social media?
  - Why not study interactions directly?
- Natural fit with interactions
  - Users only interact with small subset of others
- Degrees of influence beyond friends
  - Obesity [Fowler and Christakis, NE J. Med. 2007]
  - Altruism [Fowler et al., Econ. Let. 2009]
- Example: Zachary's Karate Club
  - Can predict behavior with network view
    - [Zachary, J. Anth. Res. 1972]



What sort of questions are we asking?

12

- Already know lots about networks
  - Scale-free [Barabasi and Albert, 1999], High clustering [Watts and Strogatz, 1998], Navigable [Adamic and Adar, 2003] [Liben-Nowell 2005], Hubs and authorities [Page and Brin, 1998] [Kleinberg, 1999], Dense core [Mislove et al. 2007]
- And have lots of models
  - Preferential Attachment [Barabasi and Albert, Nature 1999], Small-world [Watts and Strogatz, 1998], Copying [Kleinberg et al., 1999], Congestion [Mihail et al., 2003], Bowtie [Broder et al., 2000], Jellyfish [Tauro et al., 2001]
- Thus, going to focus on social aspects
  - Why do they look the way they do?
  - What can this tell us?

Outline

13

- Two parts:
- 1 Primer on social sciences
- 2 Leveraging social media

Goals

14

- Provide an overview of research on social media and networks
- Get you excited about this research area
- Give pointers to further reading
  - Papers cited throughout talk
- Spark discussion
  - Interrupt and ask questions!

PRIMER ON SOCIAL SCIENCES

Three classic papers

16

- Discuss results at a high level
  - Goal is not an in-depth discussion
- Results will frame our discussion of social media

- The Strength of Weak Ties
  - by Mark S. Granovetter
    - [American Journal of Sociology, vol. 78 issue 6. May 1973]

"Classical" sociology

18

- Focused on two topics
  - Micro-level interactions within a small group
  - Macro-level patterns within a society
- "Strong" ties considered the important ones
  - Close friends, family
  - "Weak" ties considered less important

- But, mapping not understood
  - How do large-scale patterns emerge?
  - ...certain analogies to physics...

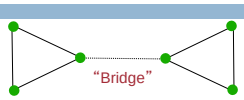
Granovetter's idea

19

- Construct simple model:
  - If two people have a common strong tie, they must have a tie between each other
- Matches intuition from real world
  - If you have two close friends, they (at least) know each other
- What are the implications of this model?

Bridges

20



- Social networks can be divided into communities
  - Clubs, schools, employers, ...
- Define a bridge as a link that is the only path between two users
- Claim: With Granovetter's assumption, bridges must be weak
  - Why?

Importance of bridges

21

- Bridges connect communities
  - Build up society from a set of communities
- Thus, weak ties (bridges) can help the micro → macro mapping
- Bridges must necessarily carry any new information
  - Example: People often find new jobs via weak ties
  - Societies with weak ties better able to adapt
  - Hence, the strength of weak ties
- But, what is the structure of weak ties at scale?
  - Are they really necessary for conveying information

- An Experimental Study of the Small World Problem
  - by Jeffery Travers and Stanley M
    - [Sociometry, vol.32 no. 4. 1969]


An Experimental Study of the Small World Problem

JEFFERY TRAVERS

Harvard University

and

STANLEY MILGRAM

The City University of New York

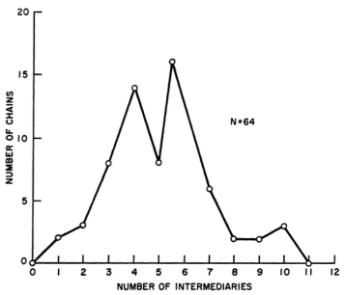
Procedure

23

- Selected 296 people in Nebraska and Boston
  - Mailed a packet containing instructions
- Packet specified a destination person
  - Name, address, profession, and city
- Asked to forward to someone known personally
  - Send a card back to Milgram
  - And add name to a roster
    - Why?

How long are the (successful) paths?

24



Implications

25

- Not only do short chains exist…
  - But people can find them!
  - With only local information
- Thus, social networks are navigable
  - 48% of 64 chains coalesced into 3 people
  - Important structural properties
- However, how did users "route"?
  - Did they rely on certain network properties?
  - Do shorter paths exist?

- Neocortex Size as a Constraint Group Size in Primates
  - by Robin I. M. Dunbar
    - [Journal of Human Evolution, vol. 2

Neocortex

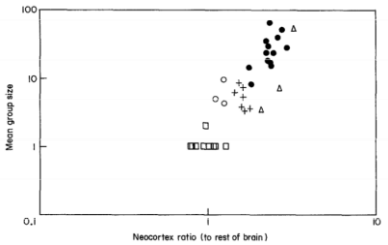
27

- Part of the brain of mammals, involved in
  - Sensory perception
  - Motor commands
  - Spatial reasoning
  - Thought and language
  - Social interactions
- Theory: large brain size due to "social" nature of primates
  - Measure "social" level by looking at typical group size
  - If true, then brain size should correlate with being "social"

Neocortex

Neocortex size and group size

28

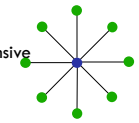


- Strong correlation observed
  - Holds across many species of primates

Implications

29

- Each individual can only maintain so many relationships
  - Bounded by neocortex size
- Not just the number of relationship, it is the intensive with which a small number of key "friendships"
  - Who likes who, who doesn't, etc
- Is this true for humans?
  - Social groups are less well-defined
  - Dunbar predicts value of 150 from neocortex size
- What about different relationship types?
  - What is the variance across individuals?



Social science primer: Summary

30

- Doing this sort of work takes significant effort!
- Key results:
  - Network structure influenced by strong/weak links
  - Networks have (navigable) short paths
  - Expected bound on degree for each node
- Do results hold for at large scale?
  - Or, for social media at all?
- What social science questions can we answer with social media?

LEVERAGING SOCIAL MEDIA

Two papers on leveraging social media

32

- Cover two topics
  - Tracking information flow
  - Applying social media to real-world problems

- Meme-tracking and the Dynam of the News Cycle
  - by Jure Leskovec, Lars Backstroı Jon Kleinberg
    - [Proceedings of KDD 2009]

Leveraging social media

34

- Networks are used to spread information
  - Can social media shed light on information flow through society?
- Focus on news media
  - How do people find out about news?
- Who "finds" stories?
  - What role does the media/social web play?
  - How do they influence each other?
- This paper: Can social media shed light on information flow?

CNN

LIVEJOURNAL

Memes

35

- Meme: Unit of culture
  - Coined by Dawkins
  - Describes evolution of culture
- Internet examples: Rickroll, LOLCat, FAIL
- Focus on memes
  - Entities (Obama) too course-grained
  - Common sequences (web 2.0) too noisy
  - Hyperlinks too fine-grained
- Use quotes to extract memes
  - "...palling around with terrorists..."

Data collected

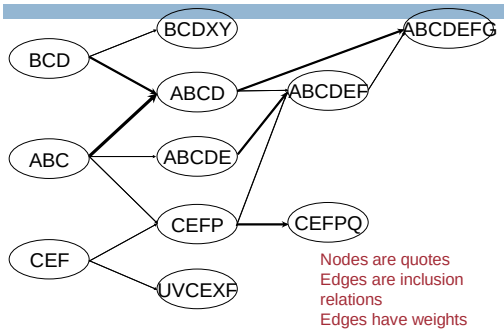
36

- Use dataset from spinn3r.com
  - August - October 2008
  - 90 million documents (blog entries/news stories)
    - 1.65 million sites
  - 112 million quotes
- Challenge: Quotes mutate
  - "...terrorists who would target their own country..."
  - "...terrorists who targeted their own country..."
  - "...terrorists who target their own country..."
  - "...terrorists who would bomb their own country..."
- How to determine which quotes are the same?

spinn3r

Clustering quotes

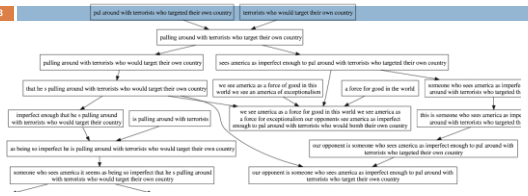
37



Nodes are quotes
Edges are inclusion relations
Edges have weights

Example of cluster

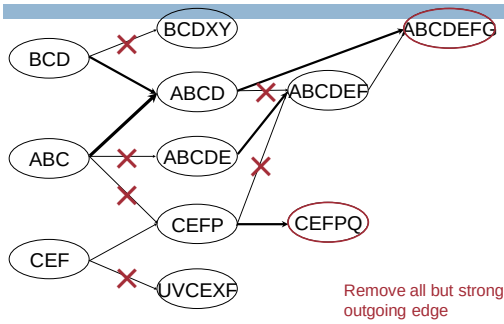
38



- All based on Sarah Palin's terrorists quote:
  - "Our opponent is someone who sees America, it seems, as being so imperfect, imperfect enough that he's palling around with terrorists who would target their own country."
- How to reduce to a single meme?

Solution: Create a DAG

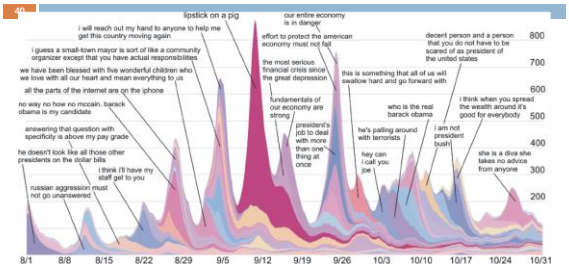
39



Remove all but strongest outgoing edge

Resulting memes

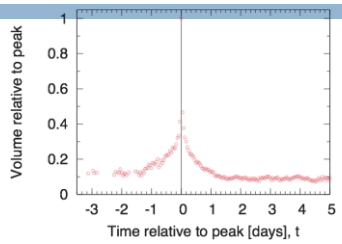
40



- Spikes show nature of 24-hour news cycle
  - Memes quickly enter and leave collective conscience

Tracking memes

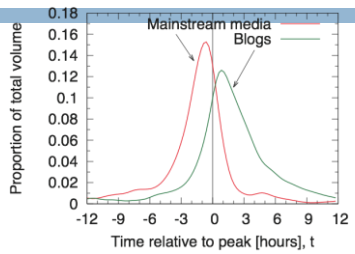
41



- First, determine "peak" intensity of each meme
  - Distinct peak present

Where do the memes come from?

42



- Second, track where articles come from
  - Media peak is 2.5 hours before blog peak
  - Blog volume persists much longer

Summary

43

- Can social media shed light on information flow?
- Collected data on over 90 million documents
  - Unprecedented scale
- Found interesting interaction between media and blogs
  - Media has short attention span
    - But causes peak intensity
  - Blogs have more persistent volume

- Predicting the Future With Social Media
  - by Sitaram Asur and Bernardo A. Huberman[Arxiv 1003.2699]

Social media and communication

45

- Social media enables communication
  - Facebook wall
  - Orkut scraps
  - Twitter tweets
- Essentially, we have microphone above the world
  - Have complete conversations for huge group of users
  - Can access collective wisdom
- Can we extract information from these conversations?
  - In aggregate?

This paper: twitter + movies

46

- Focus on twitter
  - Most data is publicly available
  - Messages are short
- Can we use twitter to predict the future?
- Focus on box-office returns for movies
  - Relatively short term (~3 week window/movie)
- Existing techniques to compare against
  - Gold standard is Hollywood Stock Exchange

Hollywood stock exchange (HSX)

47

- Example of a prediction market
  - Uses play money
- Can buy movie stocks
  - Each H$ = $1M US gross take
- Each movie has a listed delist date
  - 4 weeks after open, cashed out
  - Value is US gross take
- Surprisingly accurate
  - 32 of 39 Oscar nominees in 2007
  - 7 of 8 eventual winners

Can we use social media?

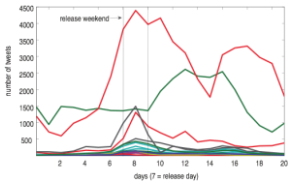
48

- Focus on mentions of 24 movies on twitter
  - Armored, Avatar, The Blind Side, The Book of Eli, Day breakers, Dear John, Did You Hear About The Morgans, Edge Of Darkness, Extraordinary Measures, From Paris With Love, The Imaginarium of Dr Parnassus, Invictus, Leap Year, LegionTwilight : New Moon, Pirate Radio, Princess And The Frog, Sherlock Holmes, Spy Next Door, The Crazies Tooth, Fairy Transylmania, When In Rome, Youth, In Revolt
- Obtained data by searching repeatedly
  - Three weeks around release date
  - Most activity in this period
  - Most money made in this period
- Total of 2.89M tweets

Making predictions

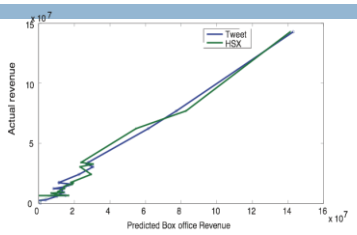
49

- Busiest time is around release
  - Promotions, advertising, ...
- Opening weekend makes most money
- Predict take by looking at pre-release tweet rate
  - How many tweets before open?
  - Compare to HSX



How accurate are the predictions?

50



- Very accurate!
  - Coefficient of determination (R^2) is 0.973
  - Versus 0.965 for HSX

Summary

51

- First look at using social media for prediction
- Relatively simple approach, naïve predictor
  - Simply looking at number of mentions before release
  - Outperformed existing gold standard
- What else can we use social media to predict?
  - Stock markets?
- But unclear causality
  - Do movie studios only promote movies they expect to be a hit?