From the Telegraph to Twitter Group Chats

James Cook

Includes work co-authored with:
From the Telegraph to Twitter Group Chats

U.S. Census via http://eh.net/encyclopedia/article/nonnenmacher.industry.telegraphic.us
Outline

• Part I: Group Chats on Twitter

• Part II: Ranking Discussion Groups

• Part III: 100 Years of News Articles
Part I: Group Chats on Twitter

Krishnaram Kenthapadi   Nina Mishra

Microsoft Research
Group Chats
#MTOS hosted by @NitrateDiva in one hour. the topic is suspense: http://t.co/8bvRl6wd
1. How do you define suspense in the cinema? As a viewer, do you consider suspense a desirable trait in a film?

2A. What is the greatest “suspense film” you’ve ever seen? Why?

2B. What’s the best, most suspenseful movie scene or sequence you can think of?
nitratediva  The Nitrate Diva
2A. What is the greatest “suspense film” you’ve ever seen? Why? #MTOS

jimsfilmmodules  James Aston
2A: Harakiri (Kobayashi’s version)—an ending that ranks amongst 1 of best made all the more memorable by the growing tension throughout #MTOS

kevrockcity  Kevin Koehler
Vertigo because it’s perfect. RT @NitrateDiva 2A. What is the greatest “suspense film” you’ve ever seen? Why? #MTOS

movietos  #MTOS
Thank you everyone. Next week’s #MTOS will have host @Thompson_film with the topic Film Noir. Do follow him and spread the word. Cheers!
nitratediva The Nitrate Diva
2A. What is the greatest “suspense film” you’ve ever seen? Why? #MTOS

jimsfilmmodules James Aston
2A: Harakiri (Kobayashi’s version) - an ending that ranks amongst 1 of best made all the more memorable by the growing tension throughout #MTOS

kevrockcity Kevin Koehler
Vertigo because it’s perfect. RT @NitrateDiva 2A. What is the greatest “suspense film” you’ve ever seen? Why? #MTOS

Next week’s #MTOS will have host @Thompson_film with the topic Film Noir.

movietos #MTOS
Thank you everyone. Next week’s #MTOS will have host @Thompson_film with the topic Film Noir. Do follow him and spread the word. Cheers!
And many more...
Group Chats
Previous Group Definitions

“. . . a collection of individuals who have relations to one another.” [Cartwright, Zander 1968]

“. . . individuals who are connected by and within social relationships.” [Greenwood 2004]

“. . . when enough people carry on public discussions long enough, with sufficient human feeling to form webs of personal relationships in cyberspace” [Rheingold 1993]
Group Chat Definition

A collection of meetings that are

- **Periodic** once per week
- **Synchronized** Sundays 2-3pm
- **Cohesive** members interact
First Attempt

A group is a collection of meetings that are periodic.
Second Attempt

A group is a collection of meetings that are periodic and **synchronized**.

One week of #MTOS

One week of #monday

Weekly TV shows
Final Definition

- A group is a collection of meetings that are periodic, synchronized and **cohesive**.

Top three #AlzChat users

Top three #monday users

😊
The Algorithm

All Hashtags

Periodic Hashtags

Periodic, Synchronized Hashtags

Periodic, Synchronized Cohesive Hashtags
Periodic

AUTOPERIOD [Vlachos, Yu, Castelli 2005]
Synchronized

One week of #MTOS

One week of #monday
Cohesive

Top three #AlzChat users

Top three #monday users
Analysis

Theorem (completeness)

The algorithm accepts hashtags that are close to being periodic, synchronized and cohesive.
Analysis

Theorem (*soundness*)

The algorithm rejects:

- Not periodic
- Not synchronized
- Not cohesive
Analysis

Theorem (*soundness*)

The algorithm rejects:

- Not periodic
- Not synchronized
- Not cohesive
Analysis

Theorem (*soundness*)

The algorithm rejects:

- Not periodic
- Not synchronized
- Not cohesive
What We Found

- Input: 2+ years of English tweets
What We Found

- Input: 2+ years of English tweets
- 1400 groups, 2.3M users
On the rise...

- Weekly groups
Part II: Ranking Discussion Groups

Krishnaram Kenthapadi    Nina Mishra    Abhimanyu Das
Sprockets

#sprocketChat
#talkSprockets
#sprockz
Sprockets

1400 Chats
27,000 Chats

# talkSprockets
# sprockz
# sprocketChat

@ alice
@ bob
@ carol
Stationary Distribution:

Pr[#talkSprockets] = 0.3

Pr[#sprockz] = 0.2

Pr[#sprocketChat] = 0.5

Final Ranking:

#sprocketChat
#talkSprockets
#sprockz
Group Preference Model

\[ \sum_{j} A_{ij} P_{jk} = \lambda D_{k} + (1 - \lambda) \]

DISCLAIMER: Not a model of reality. Use only for ranking.
Group Preference Model

Hubs and Authorities

Random Surfer Model (PageRank)
Does It Work?

Is this at all reasonable?

Theorem

If we increase one user's preference for group A (at the expense of other groups) then A's rank will not go down.

[Chien, Dwork, Kumar, Simon, Sivakumar 2003]
Comparing to the Naïve Approach

# bicycleChat
# talkSprockets
# sprockz
# sprocketChat
Experimental Setup

One Year of Tweets

27K Hashtags
Experimental Setup

2000 Test Queries

Noun Phrases (27 Million)

“someone”

“next week”

Yahoo! Groups Queries (five months)
Experimental Setup

Evaluation

“Experts” — Query appears in profile text

2000 600 Queries

Poor Quality
Experimental Setup

Evaluation

“Experts”
1. #
2. #
3. #

Algorithm 1
1. #
2. #
3. #

Algorithm 2
1. #
2. #
3. #
Experimental Setup

Evaluation

3

2

3

0
Choosing Algorithm Parameters

Teleport Probability $\lambda$

biased $>$ uniform
Choosing Algorithm Parameters

Authority Score

Weighted MAP

- # tweets with query: 0.309
- # @-mentions with query: 0.332
- # followers: 0.330
- uniform: 0.340
Choosing Algorithm Parameters

Preference Score
## Results

### Group Preference Model

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td># distinct users</td>
<td>0.168</td>
</tr>
<tr>
<td># tweets</td>
<td>0.217</td>
</tr>
<tr>
<td>Fraction of tweets with query</td>
<td>0.246</td>
</tr>
<tr>
<td>“Experts”</td>
<td>0.446</td>
</tr>
</tbody>
</table>

Weighted MAP: 0.309
Future Directions: Parts I and II

- Participation
- Research on Groups
- Knowledgable Users
- Types of query?
- Other Features?
Part III: Your Two Weeks of Fame and your Grandmother's

Alex Fabrikant     Andrew Tomkins     Atish Das Sarma

Research at Google     (eBay™)
“CNN is widely credited with initiating the acceleration of the modern news cycle with the fall 2006 debut of its spin-off channel CNN:24, which provides a breaking news story, an update on that story, and a news recap all within 24 seconds.”

- The Onion
Can we measure changes in the public's attention span?

Data source: 100 years of news
Outline

• Working with the news archive
• Measuring public attention
• Results
It's getting easier to communicate.

U.S. Census via http://eh.net/encyclopedia/article/nonnenmacher.industry.telegraphic.us
Google's News Archive

• > 60 million articles

• Substantial daily volume from 1885 to 2011.
Measuring Public Attention

2012 Jakarta Post:

'Gangnam Style' wins top MTV Europe award

Gangnam Style" phenomenon continued to forge ahead as Psy yet again added more notches to his global superstardom belt. The ubiquitous smash hit won the Best Video award at the 2012 MTV

1909 Youngstown Vindicator:

News articles have always been about people.

DAUGHTERS AT TEA TODAY

In Preparation For Next Week’s Contest Over Presidency.

Washington, April 17.—One of the chief events preceding the congress of the Daughters of the American Revolution, which opens Monday, is the tea to be given this afternoon by Mrs. James S. Sherman, wife of the vice president.
Measuring Public Attention

- Measure how long personal names stay in the news.

*Timeline for Marilyn Monroe*

*photo: Life Magazine*
Measuring Public Attention

First attempt
- Fame begins: first mention in any article
- Fame ends: last mention in any article
Three Rules

- Continuous stretches of attention
- Count each occurrence
- Normalize
A Name's Period of Fame

- Method 1: one news story
- Method 2: continuous public interest

Timeline for Marilyn Monroe
photo: Life Magazine
Results

• The median duration of fame is one week for the entire period of study (1895-2011).

• Blogger posts from 2000-2010: exactly the same result
It's getting easier to communicate.

U.S. Census via http://eh.net/encyclopedia/article/nonnenmacher.industry.telegraphic.us
It's getting easier to communicate.

U.S. Census via http://eh.net/encyclopedia/article/nonnenmacher.industry.telegraphic.us
Results

- What happens when we focus on the most famous names?
  - If we look at the 99th percentile of duration instead of the median, then we see an increasing trend since the 1940s. (left)
  - The same thing happens if we look at the 1000 most-mentioned names in each year. (right)
Future Work: Part III

- Underlying causes?
- Beyond names
- Beyond time

Culturomics!
Thanks