Ranking Twitter Discussion Groups

James Cook

Abhimanyu Das

Krishnaram Kenthapadi

Nina Mishra

COSN 2014
Outline

• Twitter discussion groups
• Our algorithm
• Theoretical results
• Evaluation
Group Chats on Twitter
[C, Kenthapadi, Mishra 2013]
MOVIE TALK ON SUNDAY

#MTOS
#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 @NitateDiva 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!
StephenS
@StephenSType1

It's that time for me. Closing: I'd be a very unhappy person without this hour every week, topic or not. Thanks @SweeterCherise #dsma

Elaine Cook
@vivizaraz

Social media enables me to access a support network, both those I know in RL and online contacts. Having child with ASD v isolating #mhchat
Find group discussions about: movies

1. #MTOS
2. #FilmCurious
3. #DriveInMob

Sort by...

# tweets with “movie”?
Fraction of tweets with “movie”?
# users who tweet “movie”?
Related Work

• Group Chats on Twitter
  [CKM 2013]

  *Algorithms for finding group chats*

  This work: *Ranking*
Related Work

- Group Chats on Twitter
  [CKM 2013]

- Search in Online Forums
  [Elsas, Carbonell 2009] [Cong et al. 2008]

Finding forum threads

This work: Finding discussion groups.
Related Work

- Group Chats on Twitter
  [CKM 2013]

- Search in Online Forums
  [Elsas, Carbonell 2009] [Cong et al. 2008]

Sprockets

# talkSprockets

# sprockz

# sprocketChat

@ alice

@ bob

@ carol
Stationary Distribution:

\[ \text{Pr}[\#\text{talkSprockets}] = 0.3 \]
\[ \text{Pr}[\#\text{sprockz}] = 0.2 \]
\[ \text{Pr}[\#\text{sprocketChat}] = 0.5 \]

Final Ranking:

\#sprocketChat
\#talkSprockets
\#sprockz
\[ M_{gh} = \begin{pmatrix} 1 \end{pmatrix} \]

\[ \sum_u A_{gu} P_{guh} \]

**Authority Scores** \( A_{gu} \)

**Preference Scores** \( P_{guh} \)

# talkSprockets

# sprockz

# sprocketChat

@ alice

@ bob

@ carol

0.1

0.2

0.3

0.7

0.1

0.6
$M_{gh} = \lambda D_h + (1 - \lambda) \sum_u A_{gu} P_{guh}$
Group Preference Model

\[ M_{gh} = \lambda D_h + (1 - \lambda) \sum_u A_{gu} P_{guh} \]

Find stationary distribution \( \pi \)

Rank \( g > h \) if \( \pi_g > \pi_h \)

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

Random Surfer Model (PageRank)

Hubs and Authorities
Stability
Stability

- PageRank and HITS are unstable.
- Our algorithm is also unstable.
Stability

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]
Rank by # times query occurs?

dementia
Rank by # times query occurs?

- dementia
- small group focused on dementia
- big news hashtag
- dementia mentioned incidentally
Example

**Theorem:** Dementia chat ranked at top.*

*(Assuming the teleport distribution is uniform.)*
Evaluation

Baseline algorithms:
• # tweets with query
• Fraction of tweets with query
• # users who tweet with query
Evaluation

- Queries
- Ground Truth
- Dataset of group discussions
Evaluation: Dataset

One year of tweets

Require at least 10 meetings

27K group discussions

One week of #MTOS
Evaluation: Queries

2000 Test Queries

Noun Phrases
(27 Million)

"someone"

"next week"

Yahoo! Groups Queries
(five months)
Evaluation: Ground Truth

“Experts” — Query appears in profile text

2000 600 Queries  Poor Quality
Evaluation: Ground Truth

“Experts”
1. #
2. #
3. #

Algorithm 1
1. #
2. #
3. #

Algorithm 2
1. #
2. #
3. #
Evaluation: Ground Truth

Evaluate by hand: 600 50 queries
### Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision@5</th>
<th>Recall@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Preference Model</td>
<td>0.40</td>
<td>0.49</td>
</tr>
<tr>
<td># distinct users</td>
<td>0.24</td>
<td>0.28</td>
</tr>
<tr>
<td># tweets</td>
<td>0.31</td>
<td>0.36</td>
</tr>
<tr>
<td>Fraction of tweets with query</td>
<td>0.27</td>
<td>0.38</td>
</tr>
<tr>
<td>(“Experts”)</td>
<td>(0.53)</td>
<td>(0.71)</td>
</tr>
</tbody>
</table>
Computing Authority Scores

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision@5</th>
<th>Recall@5</th>
</tr>
</thead>
<tbody>
<tr>
<td># tweets with query</td>
<td>0.40</td>
<td>0.49</td>
</tr>
<tr>
<td># @-mentions with query</td>
<td>0.38</td>
<td>0.47</td>
</tr>
<tr>
<td># followers</td>
<td>0.38</td>
<td>0.47</td>
</tr>
<tr>
<td>uniform</td>
<td>0.40</td>
<td>0.48</td>
</tr>
</tbody>
</table>
Summary

We designed the Group Preference Model, and found good theoretical and experimental results.
Future Directions

• Which groups are easy to join?
• Different types of query
• Personalized ranking
• Groups are always changing
• Put it online!

Thanks!