Deep Twitter Diving: Exploring Topical Groups in Microblogs at Scale


IIT Kharagpur, MPI-SWS, BESU Shibpur
The Twitter Stereotype

“Twitter provides us with a wonderful platform to discuss/confront societal problems. We trend Justin Bieber instead.”

– @LaurenLeto
Outline

• Methodology – Finding Topical Groups
  – Finding Experts
  – Finding Seekers

• How Diverse are the Topical Groups?

• Topical Groups: Identity or Bond based?
What are Topical Groups?

Topical Groups = Experts + Seekers

Experts: Users with topical knowledge

Seekers: Users interested in topical knowledge

@BarackObama
Expert on Politics

@BarackObama
Seeker on Basketball
Detecting Groups: Prior Approaches

- Graph based approaches
  - Not good for detecting “Identity based groups” [1]

- Tweet or Profile based approaches
  - Profiles: not always meaningful, not vetted
  - Tweets: small, contain lot of chatter

Outline

• Methodology – Finding Topical Groups
  – Finding Experts
  – Finding Seekers
• How Diverse are the Topical Groups?
• Topical Groups: Identity or Bond based?
Twitter Lists

- Feature for organizing followings in Twitter
- Lists have a name and description
- Tweets of the members shown separately

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>News</td>
<td>News media accounts</td>
<td>NYTimes, BBCNews, WSJ, CNNBrk, CBSNews</td>
</tr>
<tr>
<td>Music</td>
<td>Musicians</td>
<td>Eminem, BritneySpears, LadyGaga, BonJovi</td>
</tr>
<tr>
<td>Politics</td>
<td>Politicians and people who talk about them</td>
<td>BarackObama, NPRPolitics, WhiteHouse, BillMaher</td>
</tr>
</tbody>
</table>
Eric Gilbert @eegilbert
I'm an assistant professor at Georgia Tech who works on social computing. I should, therefore, be better at Twitter than I am.

Kate Starbird @katestarbird
Asst. Professor of Human Centered Design & Engineering at UW. Researcher of crisis informatics and crowdwork. Aging athlete.

Michael Bernstein @msbernst
Stanford, Assistant Professor of Computer Science. Human-computer interaction, social computing and crowdsourcing.

Ed H. Chi @edchi
HCI and Social Computing Research Scientist at Google and Google Plus; former Area Manager of Augmented Social Cognition group at PARC
Eric Gilbert @eegilbert
I'm an assistant professor at Georgia Tech who works on social computing. I should, therefore, be better at Twitter than I am.

Kate Starbird @katestarbird
Asst. Professor of Human Centered Design & Engineering at UW. Researcher of crisis informatics and crowdwork. Aging athlete.

Michael Bernstein @msbernst
Stanford, Assistant Professor of Computer Science. Human-computer interaction, social computing and crowdsourcing.

Ed H. Chi @edchi
HCI and Social Computing Research Scientist at Google and Google Plus; former Area Manager of Augmented Social Cognition group at PARC
Eric Gilbert @eegilbert
I'm an assistant professor at Georgia Tech who works on social computing. I should, therefore, be better at Twitter than I am.

Kate Starbird @katestarbird
Asst. Professor of Human Centered Design & Engineering at UW. Researcher of crisis informatics and crowdwork. Aging athlete.

Michael Bernstein @msbernst
Stanford, Assistant Professor of Computer Science. Human-computer interaction, social computing and crowdsourcing.

Ed H. Chi @edchi
HCI and Social Computing Research Scientist at Google and Google Plus; former Area Manager of Augmented Social Cognition group at PARC
Eric Gilbert @eegilbert
I'm an assistant professor at Georgia Tech who works on social computing. I should, therefore, be better at Twitter than I am.

Kate Starbird @katestarbird
Asst. Professor of Human Centered Design & Engineering at UW. Researcher of crisis informatics and crowdwork. Aging athlete.

Michael Bernstein @msbernst
Stanford, Assistant Professor of Computer Science. Human-computer interaction, social computing and crowdsourcing.

Ed H. Chi @edchi
HCI and Social Computing Research Scientist at Google and Google Plus; former Area Manager of Augmented Social Cognition group at PARC.
If one is included in a number of lists, on the same topic, one is likely to be an expert on the topic.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Experts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Music</td>
<td>Lady Gaga, ColdPlay, Katy Perry, Dallas Martin [VP Warner Records]</td>
</tr>
<tr>
<td>Politics</td>
<td>Barack Obama, Al Gore, Scott Fluhr [Harrison County GOP chairman]</td>
</tr>
<tr>
<td>Forensics</td>
<td>Sans Institute, Forensic Focus, Michael Murr [Forensic Scientist]</td>
</tr>
<tr>
<td>Geology</td>
<td>GeoSociety, Kim Hannula [Geology Prof.], Garry Hayes [Geology Teacher]</td>
</tr>
</tbody>
</table>

Outline

• Methodology – Finding Topical Groups
  – Finding Experts
  – Finding Seekers

• How Diverse are the Topical Groups?

• Topical Groups: Identity or Bond based?
Kate Starbird
@katestarbird
Asst. Professor of Human Centered Design & Engineering at UW.
Researcher of crisis informatics and crowdwork. Aging athlete.
Seattle, WA

Abraham Flaxman @healthyalgo
Math/TCS applied to Global Health

Medical HumanFactors @MedicalHFE
National Center for Human Factors in Healthcare @MedStarHealth: Industrial Systems Eng|Cog Psych|Syst Safety Eng|#Simulation|#PtSafety|#HFsafety|#HumanFactors

Seattle Storm @seattlestorm

Sarah Webber @sarahwebber01
I love exploring technology and helping people put it to good use. Volunteer techie. UX consultant. BA. HCI student. Tech writer.

Mathias Klang @Klangable
Associate Professor, Research on Technology Regulation, Rights, Social Media, Privacy, e-books, Copyright & more
Marta Dydek: Retired basketball player. NBA and WNBA commentator at Canal+ Sport.

Rushia Brown: WNBA Vet. President/Founder of Women's Professional Basketball Alumni Association. Paying it Forward by molding the lives of young ladies ON & OFF the court!

Debbie Antonelli: College bball/ WNBA analyst/co-host ESPN Women's Bball podcast/mom to 3 boys/can still run post pattern, hook slide and knock down 3 pters.


WNBA: News & notes directly from the WNBA.

Mary Murphy: Basketball is my life! Cover the Pac-12, Big Ten. BTW...I like other sports too. NFL, MLB, NBA...follow them all.

Evan: Assistant Coach at USC! Basketball has taken me on an incredible ride at Santa Clara, UCSB, Stanford and Missouri! Love the journey!

Stanford W Basketball: The official Twitter account for the two-time National Champion Stanford Women's Basketball team. #gostanford
Marta Dydek @mdydek
Retired basketball player. NBA and WNBA commentator at Canal+Sport.

Rushia Brown @RushiaB
WNBA Vet. President/Founder of Women's Professional Basketball Alumni Association. Paying it forward by molding the lives of young ladies ON & OFF the court!

Debbie Antonelli @debbieantonelli
College bbball/WNBA analyst/co-host ESPN Women’s Bball podcast/mom to 3 boys/can still run post pattern, hook slide and knock down 3 pters

Chantelle @MissChantelle
Jesus saves! Disciple. Former WNBA. Vandy WBB Alum. Corporate America. Call Me Miss. John 1:5. Check out my blog

Mary Murphy @realmarymurray
Basketball is my life! Cover the Pac-12, Big Ten. BTW...I like other sports too. NFL, MLB, NBA...follow them all.

Evan @USCCoachEvan
Assistant Coach at USC! Basketball has taken me on an incredible ride at Santa Clara, UCSB, Stanford and Missouri! Love the journey!

Stanford WBBasketball @StanfordWBB
The official Twitter account for the two-time National Champion Stanford Women's Basketball team. #gostanford
Marta Dydek @mdydek
Retired basketball player. NBA and WNBA commentator at Canal+ Sport.

Rushia Brown @RushiaB
WNBA Vet. President/Founder of Women’s Professional Basketball Alumni Association. Paying it Forward by molding the lives of young ladies ON & OFF the court!

Debbie Antonelli @debbieantonelli
College bball/WNBA analyst/co-host ESPN Women’s Bball podcast/mom to 3 boys/can still run post pattern, hook slide and knock down 3 pters

Chantelle @MissChantelle
Jesus saves! Disciple. Former WNBA. Vandy WBB Alum. Corporate America. Call Me Miss. John 1:5. Check out my blog

Mary Murphy @realmarymurphy
Basketball is my life! Cover the Pac-12, Big Ten. BTW...I like other sports too. NFL, MLB, NBA...follow them all.

Evan @USCCoachEvan
Assistant Coach at USC! Basketball has taken me on an incredible ride at Santa Clara, UCSB, Stanford and Missouri! Love the journey!

Stanford WBasketball @StanfordWBB
The official Twitter account for the two-time National Champion Stanford Women’s Basketball team. #gostanford
If one is following many experts, on the same topic, one is likely to be interested in the topic.
Who likes What - Infer Topical interests of Twitter Users

Kate Starbird
@katestarbird

Asst. Professor of Human Centered Design & Engineering at UW. Researcher of crisis informatics and crowdwork. Aging athlete.

Seattle, WA

Interests of: Kate Starbird

news politics
earthquake
seattle area
ems
academic

ict4d
researchers

new media
govt

disaster
response

preparedness
geospatial

humanitarian
mapping

hci
wba
public safety

Stanford revolution
influential mena
Who likes What - Infer Topical interests of Twitter Users

Kate Starbird
@katestarbird
Asst. Professor of Human Centered Design & Engineering at UW. Researcher of crisis informatics and crowdwork. Aging athlete.
Seattle, WA

Interests of: Kate Starbird

WNBA
Who likes What - Infer Topical interests of Twitter Users

critical play
@criticalplay

Patrick Meier
@PatrickMeier
Topical Groups

Topical Group = Experts + Seekers

Experts and Seeker sets overlap.
Outline

- Methodology – Finding Topical Groups
  - Finding Experts
  - Finding Seekers
- How Diverse are the Topical Groups?
- Topical Groups: Identity or Bond based?
Scalability of our Approach

- First 38 Million users in Twitter
- 88 Million lists. 1.5 Billion links
- 36 Thousand Topical Groups
- Covering 49.5% users
- Covering 94.3% links
# Diversity: Topics and Group Size

<table>
<thead>
<tr>
<th>No. of seekers</th>
<th>Number of experts</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 1K</td>
<td>geology, karate, malaria, neurology, tsunami, psychiatry, radiology, pediatrics, dermatology, dentistry, volleyball, philosophers, tarot, perfume, florists, copywriters, taxi, esperanto</td>
</tr>
<tr>
<td>1K – 5K</td>
<td>biology, chemistry, swimmers, astrophysics, multimedia, semiconductor, renewable-energy, breast-cancer, judaism, astrology, sociology, geography, forensics, anthropology, genealogy, archaeology, gluten, diabetes, neuroscience, architects, insurance, second-life, police, progressives, creativity</td>
</tr>
<tr>
<td>5K – 10K</td>
<td>malware, gnu, robot, chicago-sports, gospel-music, space-exploration, wall-street, horror, agriculture, atheism, attorneys, furniture, art-galleries, ubuntu, psychology, poetry, catholic, hospitals, autism, jazz</td>
</tr>
<tr>
<td>10K – 50K</td>
<td>ipod, ipad, virus, Liverpool-FC, choreographers, heavy-metal, backstreet-boys, world-cup, olympics, physics, theology, earthquake, opera, makeup, Adobe, wrestlers, typography, american-idol, tennis, linux, astronomy, yoga, animation, manga, doctors, realtors, wildlife, rugby, forex, php, java, law, history, beer, golf, librarians, theatre, military, poker, conservatives, vegan</td>
</tr>
<tr>
<td>50K– 100K</td>
<td>bbc-radio, UK-celebs, christian-leaders, superstars, hackers, programmers, bicycle, GOP, fantasy-football, NCAA, wwe, sci-fi, medicine, cyclists, investors, recipes, NHL, xbox, triathlon, Google, hotels, museums, hockey, architecture, charities, weather, space</td>
</tr>
<tr>
<td>&gt; 100K</td>
<td>headlines, brits, pop-culture, gospel, BBC, reality-tv, bollywood, religion, actresses, gadgets, graphic-design, directors, lifestyle, gossip, commentators, youtube, books, government, comedy, environment, baseball, soccer, hollywood, iphone, economics, money</td>
</tr>
<tr>
<td></td>
<td>fashion, education, wine, photography, radio, restaurants, science, SEO, music, tech, business, politics, food, sports, celebs, health, media, bloggers, travel, writers</td>
</tr>
</tbody>
</table>
# A Small Number of Very Popular Groups

<table>
<thead>
<tr>
<th>No. of seekers</th>
<th>Number of experts</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 1K</td>
<td>(5416) geology, karate, malaria, neurology, tsunami, psychiatry, radiology, pediatrics, dermatology</td>
</tr>
<tr>
<td>1K – 5K</td>
<td>(915) istry, astrophysics, media, renewable energy, breast-cancer</td>
</tr>
<tr>
<td>5K – 10K</td>
<td>(166) robotics, gospel-exploration</td>
</tr>
<tr>
<td>10K – 50K</td>
<td>(174) virus, choreography, metal, world-cup</td>
</tr>
<tr>
<td>50K – 100K</td>
<td>(7) bicycle, celebs, leaders, superstars</td>
</tr>
<tr>
<td>&gt; 100K</td>
<td>(3) headlines, brits</td>
</tr>
<tr>
<td></td>
<td>(58) religion, actresses, gadgets, graphic-design, directors, lifestyle, gossip, commentators, youtube</td>
</tr>
</tbody>
</table>

- **37** hotels, museums, hockey, architecture, charities, weather, space
- **140** books, government, comedy, environment, baseball, soccer, hollywood, iphone, economics, money
- **25** fashion, education, wine, photography, radio, restaurants, science, SEO
- **17** music, tech, business, politics, food, sports, celebs, health, media, bloggers, travel, writers
### Thousands of Specialized Niche Groups

<table>
<thead>
<tr>
<th>No. of seekers</th>
<th>Number of experts</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 1K</td>
<td></td>
</tr>
<tr>
<td>(5416) geology, karate, malaria, neurology, tsunami, psychiatry, radiology, pediatrics, dermatology, dentistry</td>
<td>(132) volleyball, philosophers, tarot, perfume, florists, copywriters, taxi, esperanto</td>
</tr>
<tr>
<td>1K – 5K</td>
<td></td>
</tr>
<tr>
<td>(915) biology, chemistry, swimmers, astrophysics, multimedia, semiconductor, renewable-energy, breast-cancer, judaism</td>
<td>(428) painters, astrology, sociology, geography, forensics, anthropology, genealogy, archaeology, gluten, diabetes, neuroscience</td>
</tr>
<tr>
<td>5K – 10K</td>
<td></td>
</tr>
<tr>
<td>(166) malware, robot, chicago, gospel-music, exploration, world-cup, Ncaa, wwe, sci-fi</td>
<td>(58) religion, actresses, gadgets, graphic-design, directors, lifestyle, gossip, commentators, youtube</td>
</tr>
<tr>
<td>10K – 50K</td>
<td></td>
</tr>
<tr>
<td>(174) ipod, virus, liverpool, choreographers, metal, backstreet, world-cup, (49) pop-culture, gospel, bbc, reality-tv, bollywood</td>
<td>(140) books, government, comedy, environment, baseball, soccer, hollywood, iphone, economics, money</td>
</tr>
<tr>
<td>50K – 100K</td>
<td></td>
</tr>
<tr>
<td>(7) bbc-radio, celebs, cch, leaders, superheroes, ties, weather, space</td>
<td>(25) fashion, education, wine, photography, radio, restaurants, science, seo</td>
</tr>
<tr>
<td>&gt; 100K</td>
<td></td>
</tr>
<tr>
<td>(3) headlines, brits</td>
<td>(17) music, tech, business, politics, food, sports, celebs, health, media, bloggers, travel, writers</td>
</tr>
</tbody>
</table>
The Twitter Stereotype

popular news, celebrities, current events, and chatter

- “What is Twitter”, Kwak et. al., WWW 2010
- “Who says What to Whom on Twitter”, Wu et. al., WWW 2011
Breaking the Stereotype

• Exploring Topical Groups at Scale

• Groups Include
  – Politics, music, ...
  – Geology, neurology, karate, malaria, astrophysics, renewable energy, judaism, forensics, genealogy, esperanto, ...
Outline

• Methodology – Finding Topical Groups
  – Finding Experts
  – Finding Seekers

• How Diverse are the Topical Groups?

• Topical Groups: Identity or Bond based?
Why do groups and communities form?

“Common Identity and Bond Theory”

Identity Based Groups:
Sports Fans
Identity Based Groups:
Professional Groups e.g. CSCW
Bond Based Groups:
Family and Friends
Common Identity vs. Common Bond Theory

Identity Based Groups
- Low Reciprocity
- Low Personal Interactions
- High Topicality

Bond Based Groups
- High Reciprocity
- High Personal Interactions
- Low Topicality
We picked 50 topical groups for detailed analysis.
The 50 groups are spread across the spectrum.
Reciprocity and Interactions

• Reciprocity in Topical Groups is Low
  – High between experts (0.3-0.6)
  – Low between experts and seekers (0.2)

• One-to-one interaction is Low
  – Further details in paper
Topicality of Discussions

http:// ...

AlchemyAPI™

Named Entities  Keywords
Expert's Tweets are very Topical

Related urls are more than 50% for 36 groups.

Implication: Useful for content mining systems.
Topical Groups are Identity Based

- **Low** Overall Reciprocity
- **Low** Personal Interactions
- **Highly** Topical Tweets

Implications: Difficult to detect via community detection
Implications

- Topical News and Search Systems
- Topical Recommender Systems
- Emerging Expert Detection Systems
Conclusion

- Twitter is a rich source of niche content
  - We found thousands of groups on niche topics

- Topical Groups are Identity Based Groups
  - With low connectivity and high topicality
Conclusion

• Twitter is a rich source of niche content
  – We found thousands of groups on niche topics

• Topical Groups are Identity Based Groups
  – With low connectivity and high topicality

Thank You!
Backup Slides
BGLL communities have much lower cut ratio and conductance.
F-Score between Topical Groups and Best Matching BGLL Groups

Topical Groups and BGLL communities don't match.
Expert URLs vs. Random URLs

For niche topics, expert URLs are 10 times more on topic.
Expert Proximity

Experts are within two hops of 60-80% other experts.
Density of Expert Mention Network

Destiny of mentions is much lower than connections.
Mashable Lists
**Tech** by Mashable
Experts and sources to keep up with the latest in tech.
69 members

Tim O'Reilly @timoreilly
Founder and CEO, O'Reilly Media. Watching the alpha geeks, sharing their stories, helping the future unfold.

Cory Doctorow @doctorow
Writer, blogger, activist. If you want a reply, use email. Blog suggestions here: [http://www.boingboing.net/suggest.html](http://www.boingboing.net/suggest.html)

Guy Kawasaki @GuyKawasaki
Advises Motorola. Author of APE: Author, Publisher, Entrepreneur. Free Kindle version here
Former chief evangelist of Apple.
Tim O'Reilly @timoreilly
Founder and CEO, O'Reilly Media. Watching the alpha geeks, sharing their stories, helping the future unfold.

Cory Doctorow @doctorow
Writer, blogger, activist. If you want a reply, use email. Blog suggestions here: http://www.boingboing.net/suggest.html

Guy Kawasaki @GuyKawasaki
Former chief evangelist of Apple.
Tech by Mashable
Experts and sources to keep up with the latest in tech.
69 members

Tim O'Reilly @timoreilly
Founder and CEO, O'Reilly Media. Watching the alpha geeks, sharing their stories, helping the future unfold.

Cory Doctorow @doctorow
Writer, blogger, activist. If you want a reply, use email. Blog suggestions here: http://www.boingboing.net/suggest.html

Guy Kawasaki @GuyKawasaki
Former chief evangelist of Apple.
Tech by Mashable
Experts and sources to keep up with the latest in tech.
69 members

Tim O'Reilly @timoreilly
Founder and CEO, O'Reilly Media. Watching the alpha geeks, sharing their stories, helping the future unfold.

Cory Doctorow @doctorow
Writer, blogger, activist. If you want a reply, use email. Blog suggestions here: http://www.boingboing.net/suggest.html

Guy Kawasaki @GuyKawasaki
Former chief evangelist of Apple.
Love food? Here are chef's, cooks and others in food to follow.

14 members

News, recipes & reviews

News staff: @Russ_Parsons1 @bettyhallock @thejgold @sirenevibara @Jenn_Harris_ @noellecarter @ReneLynch

Katie Lee @KatieLeeKitchen
Co-Host of The Kitchen on Food Network, and author of Groundswell, The Comfort Table, and The Comfort Table: Recipes for Everyday Occasions

curious, inquisitive, experimental entrepreneur who cares about the world we live in
Food by Mashable
Love food? Here are chef’s, cooks and others in food to follow.
14 members

L.A. Times Food @latimesfood
News, recipes & reviews @latimes
Food staff: @Russ_Parsons1
@bettyhallock @thejgold
@sirenevirbila @Jenn_Harris_
@noellecarter @ReneLynch

Katie Lee @KatieLeeKitchen
Co-Host of The Kitchen on Food Network, and author of Groundswell, The Comfort Table, and The Comfort Table: Recipes for Everyday Occasions

Martha Stewart @MarthaStewart
curious, inquisitive, experimental entrepreneur who cares about the world we live in
Food by Mashable
Love food? Here are chef’s, cooks and others in food to follow.
14 members

L.A. Times Food @latimesfood
News, recipes & reviews @latimes
Food staff: @Russ_Parsons1 @bettyhallock @thejgold @sirenevirbila @Jenn_Harris_ @noellecarter @ReneLynch

Katie Lee @KatieLeeKitchen
Co-Host of The Kitchen on Food Network, and author of Groundswell, The Comfort Table, and The Comfort Table: Recipes for Everyday Occasions

Martha Stewart @MarthaStewart
curious, inquisitive, experimental entrepreneur who cares about the world we live in
“Cognos: Crowdsourcing Search for Topic Experts in Microblogs”
Ghosh et. al, SIGIR, 2012.