Deep Twitter Diving:

Exploring Topical Groups in Microblogs at Scale

P. Bhattacharya, S. Ghosh, J. Kulshrestha, M. Mondal, M. B. Zafar, N. Ganguly, and K. P. Gummadi

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

[1] Grabowicz et. al., "Distinguishing topical and social groups based on common identity and bond theory", WSDM 2013

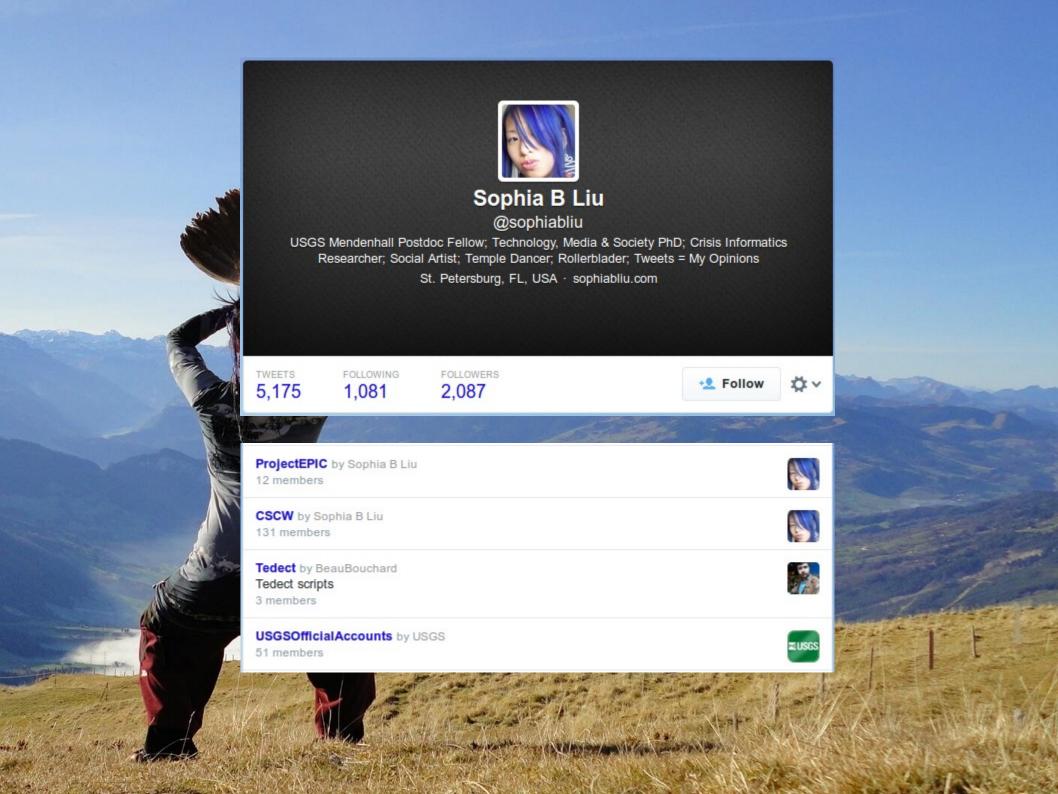
Outline

- 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

Name	Description	Members
News	News media accounts	NYTimes, BBCNews, WSJ, CNNBrk, CBSNews
Music	Musicians	Eminem, BritneySpears, LadyGaga, BonJovi
Politics	Politicians and people who talk about them	BarackObama, NPRPolitics, WhiteHouse, BillMaher





A public list by Sophia B Liu

MEMBERS

SUBSCRIBERS

131

14



@sophiabliu

USGS Mendenhall Postdoc Fellow; Technology, Media & Society PhD; Crisis Informatics Researcher; Social Artist; Temple Dancer; Rollerblader; Tweets = My Opinions

ersburg, FL, USA · sophiabliu.com



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.



E

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



51 members



A public list by Sophia B Liu

MEMBERS

SUBSCRIBERS

131

14



@sophiabliu

USGS Mendenhall Postdoc Fellow; Technology, Media & Society PhD; Crisis Informatics Researcher; Social Artist; Temple Dancer; Rollerblader; Tweets = My Opinions

rsburg, FL, USA · sophiabliu.com



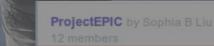
Eric Gilbert @eegilbert

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.



3

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



51 member



A public list by Sophia B Liu

MEMBERS

SUBSCRIBERS

131

14

)87





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





A public list by Sophia B Liu

MEMBERS

SUBSCRIBERS

131

14



@sophiabliu

USGS Mendenhall Postdoc Fellow; Technology, Media & Society PhD; Crisis Informatics Researcher; Social Artist; Temple Dancer; Rollerblader; Tweets = My Opinions

ersburg, FL, USA · sophiabliu.com



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.



ProjectEPIC by Sophia B Liu

12 members



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.

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

Ghosh et. al., "Cognos: Crowdsourcing search for Topic Experts in Microblogs", SIGIR 2012

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



Follow

Follow



Medical HumanFactors @MedicalHFE

National Center for Human Factors in Healthcare @MedStarHealth: Industrial

Systems EnglCog Psych|Syst Safety Eng|#Simulation|#PtSafety|#HFsafety|#Huma

nFactors



Seattle Storm @ @seattlestorm

Official Twitter of the Seattle Storm, the 2004 and 2010 WNBA champions. Owned by Force 10 Hoops. Est. 2000. instagram.com



Follow



/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.



Follow



Mathias Klang @Klangable

Associate Professor, Research on Technology Regulation, Rights, Social Media, Privacy, e-books, Copyright & more





Follow



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 &





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

nan @healthyalgo ed to Global Health

eattle Storm



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



OFF the court!

Medical HumanFactors @MedicalHFE WNBA @ @WNBA dustrial News & notes directly from the WNBA. afety



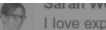
Mary Murphy @realmarymurphy Basketball is my life! Cover the Pac-12, Big Ten. BTW...I like other sports too. NFL, MLB, NBA...follow them all.



Evan @USCCoachEvan

blog

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



I love exploring technology and helping

people put it UX consultar



+ Follow

Mathias Kla Associate Pr Regulation, e-books, Copyright & more



Stanford WBasketball @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!

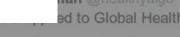
Kate @k



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

nan @healthyalgo ed to Global Health



Medical HumanFactors @MedicalHFE



Chantelle @MissChantelle

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



WNBA @ @WNBA

101

News & notes directly from the WNBA.

tors in ndustrial afety



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.

eattle Storm

Evan @USCCoachEvan

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



I love exploring technology and helping

UX consultar



• Follow

people put it



Regulation,

e-books, Copyright & more

Stanford WBasketball @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 &





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

nan @healthyalgo ed to Global Health



Medical HumanFactors @MedicalHFE



OFF the court!

WNBA @ @WNBA

News & notes directly from the WNBA.

tors in ndustrial afety safety|#Hu



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

nFactors



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!

vebber0



I love exploring technology and helping

people put it to good use Voluntor
UX consultar Stanford



Follow

Mathias Kla Associate Pr Regulation, Construction of the last of th

Stanford WBasketball @StanfordWBB

The official Twitter account for the two-time National Champion Stanford Women's Basketball team. #gostanford

e-books, Copyright a more

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



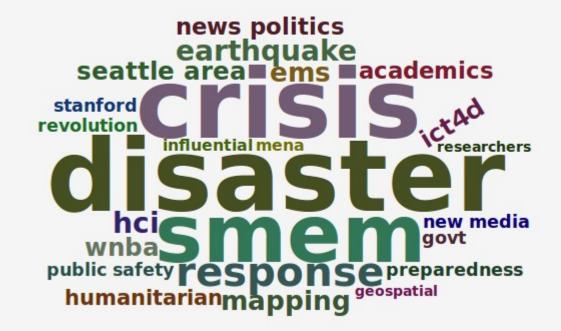
Kate Starbird

@katestarbird

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

Seattle, WA

Interests of: Kate Starbird



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



Who likes What - Infer Topical interests of Twitter Users







international development
academics global health
democracy ct4 dsmem
privacy aid ct4 dsmem
ngos
mapping crisis stechies
innovators gov20 qatardisaster
new media
humanitarian
emergency management

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

No. of	Number of experts						
seekers	< 100	100 - 500	500 – 1K 1K – 5K		5K – 10K	> 10K	
< 1K	(5416) geology, karate, malaria, neurology, tsunami, psychiatry, radiology, pediatrics, dermatology, dentistry	(132) volleyball, philosophers, tarot, perfume, florists, copywriters, taxi, esperanto					
1K – 5K	(915) biology, chemistry, swimmers, astrophysics, multimedia, semiconductor, renewable-energy, breast-cancer, judaism	(428) painters, astrology, sociology, geography, forensics, anthropology, genealogy, archaeology, gluten, diabetes, neuroscience	(17) architects, insurance, second-life, police, progressives, creativity				
5K – 10K	(166) <i>malware</i> , gnu, robot, chicago-sports, gospel-music, space-exploration, wall-street	(202) horror, agriculture, atheism, attorneys, furniture, art-galleries, ubuntu	(34) psychology, poetry, catholic, hospitals, autism, jazz	(2) coffee, dealers			
10K – 50K	(174) ipod, ipad, virus, Liverpool-FC, choreographers, heavymetal, backstreet-boys, world-cup,	(312) olympics, physics, theology, earthquake, opera, makeup, Adobe, wrestlers, typography, american-idol	(146) tennis, linux, as- tronomy, yoga, anima- tion, manga, doctors, realtors, wildlife, rugby, forex, php, java,	(67) law, history, beer, golf, librari- ans, theatre, military, poker, conservatives, vegan			
50K- 100K	(7) bbc-radio, UK-celebs, christian-leaders, superstars	(61) hackers, programmers, bicycle, GOP, fantasy-football, NCAA, wwe, sci-fi	(35) medicine, cyclists, investors, recipes, NHL, xbox, triathlon, Google	(37) hotels, mu- seums, hockey, architecture, chari- ties, weather, space			
> 100K	(3) headlines, brits	(49) pop-culture, gospel, BBC, reality-tv, bollywood	(58) religion, actresses, gadgets, graphic-design, directors, lifestyle, gossip, commentators, youtube	(140) books, govern- ment, comedy, en- vironment, baseball, soccer, hollywood, iphone, economics, money	(25) fashion, education, wine, photog- raphy, radio, restaurants, science, SEO	(17) music, tech, business, politics, food, sports, celebs, health, media, bloggers, travel, writers	

A Small Number of Very Popular Groups

No. of	of Number of experts						
seekers	< 100	100 – 500	500 – 1K	1K – 5K	5K – 10K	> 10K	
< 1K	(5416) geology, karate,	(132) volleyball,					
	malaria, neurology,	philosophers, tarot,					
	tsunami, psychiatry,	perfume, florists, copy-					
	radiology pediatrics	_				7	
177	$\frac{\text{dermate}}{\text{(37)}} \text{h}$	notels, mu-					
1K – 5K	istry, seums,	hockey,					
	1110 01019	ture, chari-					
	breast-c ties, we	ather, space					
5K – 10K	(166) (140) bo	ooks, govern-	(25) <i>fashio</i>	on, (17) mi	usic, tech,		
	explorate	comedy, en-	education,	I	s, politics,		
10K - 50K	(174) vironme	ent, baseball,	wine, photo	g- food,	sports,		
30K		hollywood,	raphy, radi	io, <i>celebs</i> ,	health,		
	metal, iphone,	economics,	restaurants,	media,	bloggers,		
50K- 100K	(7) bl money		science, SEO	travel, v	writers		
	leaders, superstars	GOP, fantasy-football, NCAA, wwe, sci-fi	xbox, triathlon, Google	architecture, charities, weather, space			
> 100K	(3) headlines, brits	(49) pop-culture, gospel, BBC, reality-tv, bollywood	(58) religion, actresses, gadgets, graphic-design, directors, lifestyle, gossip, commentators, youtube	(140) books, govern- ment, comedy, en- vironment, baseball, soccer, hollywood, iphone, economics, money	(25) fashion, education, wine, photog- raphy, radio, restaurants, science, SEO	(17) music, tech, business, politics, food, sports, celebs, health, media, bloggers, travel, writers	

Thousands of Specialized Niche Groups

No. of	Number of experts							
seekers	< 100	100 – 500	500 – 1K		1K – 5K	5K – 10K	> 10K	
< 1K	(5416) geology, karate,	(132) volleyball,						
	malaria, neurology,	philosophers, tarot,						
	tsunami, psychiatry	nerfume florists conv		(1.0.0)				
	radiology, ped (54 dermatology, der	116) geology, k	arate,	(132)	volley	ball,		
1K -	(915) biology, ma	laria, neuro	ology,	philos	ophers, t	arot,		
5K	istry, swi astrophysics, tsu	<i>nami</i> , psych	niatry.	perfur	ne, florists, c	opv-		
		* · · · · · · · · · · · · · · · · · · ·	•	•		A *		
	renewable-energy rac	liology, pedia	atrics,	writers, taxi, esperanto				
	breast-cancer, ju der	matology, dent	istry					
5K -	(100) maiware,			(420)		1 1		
10K	robot, chicago (9]	l 5) <i>biology</i> , (cnem-	(428)	painters, as	strot-		
	exploration, wall <i>ist</i>	ry, swim	mers	ogy s	ociology, geo	ogra-		
10K -	(174) inod							
50K	virus, Liverpo ast	rophysics, 1	multi-	phy, J	<i>forensics</i> , an	thro-		
	choreographers, me	media, semiconductor,		pology, genealogy, ar-				
	mean, backstree	KSHOU						
50K-	world-cup, renewable-energy, chaeology, gluten, dia-							
100K	(7) bbc-radio,	(7) 000-14410,						
1001	leaders, superstate	ast-cancer, juda	118111	betes,	Heuroscience			
		NCAA, wwe, sci-fi		,	ties, weather, space			
>	(3) headlines, brits	(49) pop-culture,	(58) religion		(140) books, govern-	(25) fashion,	(17) music, tech,	
100K		gospel, BBC, reality-tv,	gadgets,	U 1	ment, comedy, en-	education,	business, politics,	
		bollywood	design,	directors,	vironment, baseball,	wine, photog-	food, sports,	
			lifestyle, go		soccer, hollywood,	raphy, radio,	celebs, health,	
	mentators, y		outube	iphone, economics, money	restaurants, science, SEO	media, bloggers, travel, writers		
					Honey	science, BEO	uavei, willers	

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"

Prentice et. al. "Asymmetries in Attachments to Groups and to Their Members: Distinguishing Between Common-Identity and Common-Bond Groups", Personality and Social Psychology Bulletin, 1994

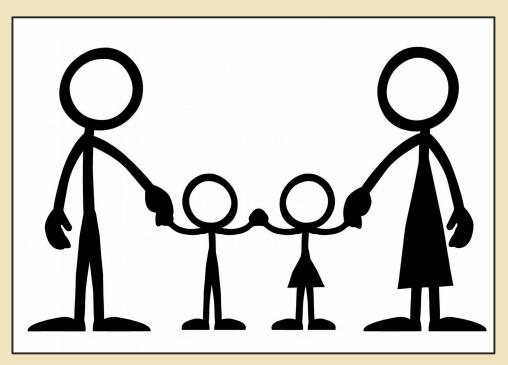
Identity Based Groups: Sports Fans



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

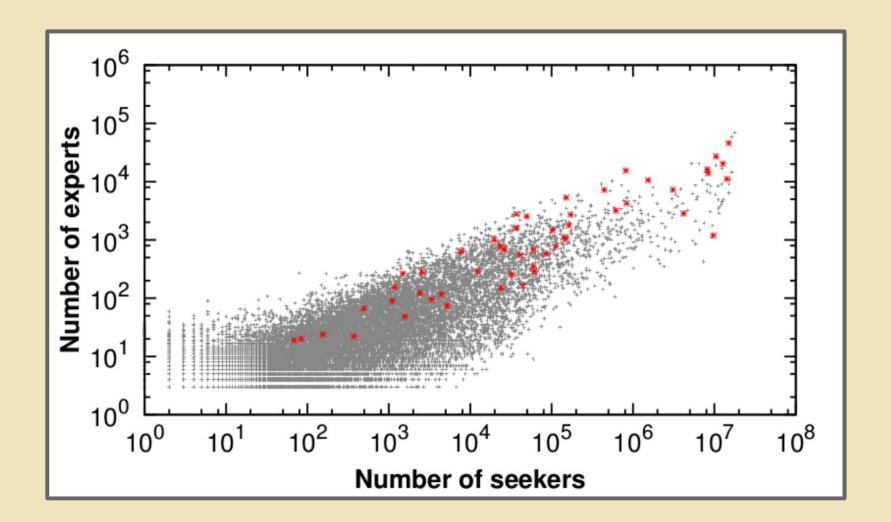
High Topicality

Bond Based Groups

High Reciprocity

High Personal Interactions

Low Topicality

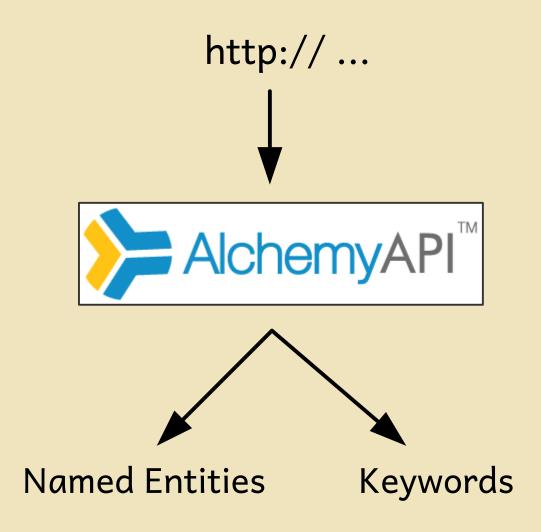


We picked 50 topical groups for detailed analysis The 50 groups are spread across the spectrum

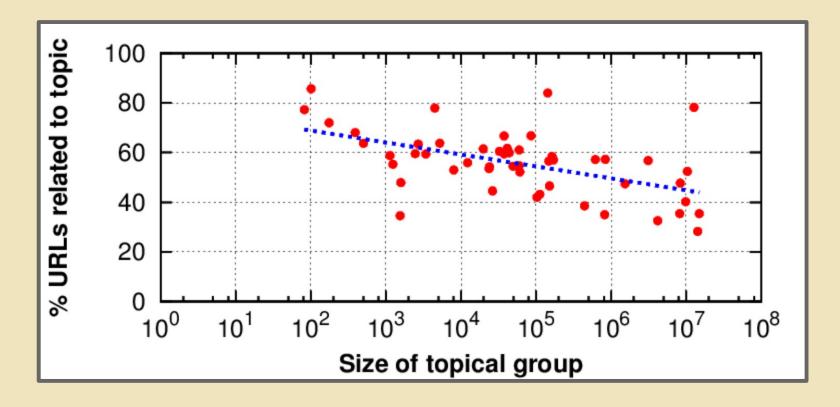
Reciprocity and Interactions

- 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



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

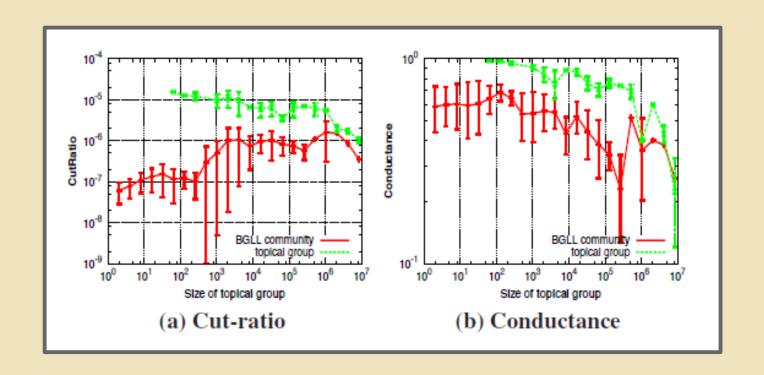
- Twitter is a rich source of niche content
 - We found thousands of groups on niche topics

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

Thank You!

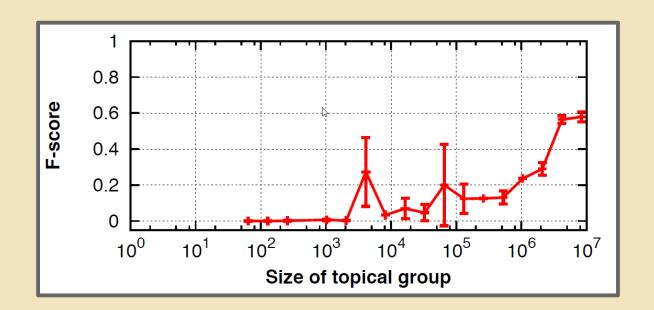
Backup Slides

Cut Ratio and Conductance of Topical Groups



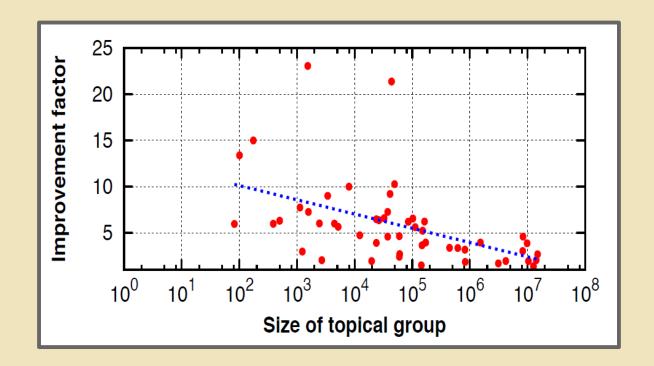
BGLL communities have much lower cut ratio and conductance.

F-Score between Topical Groups and Best Matching BGLL Groups



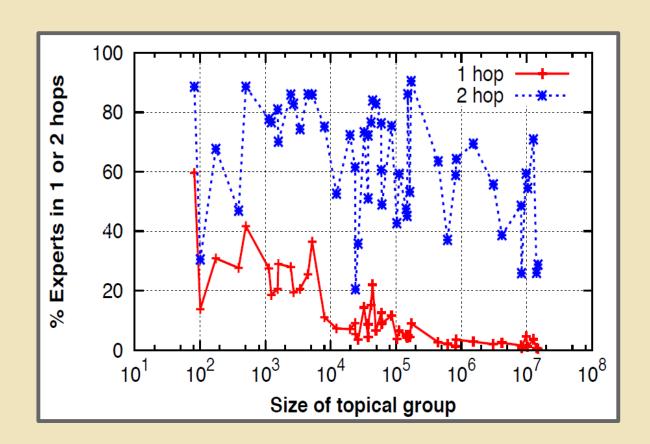
Topical Groups and BGLL communities don't match.

Expert URLs vs. Random URLs



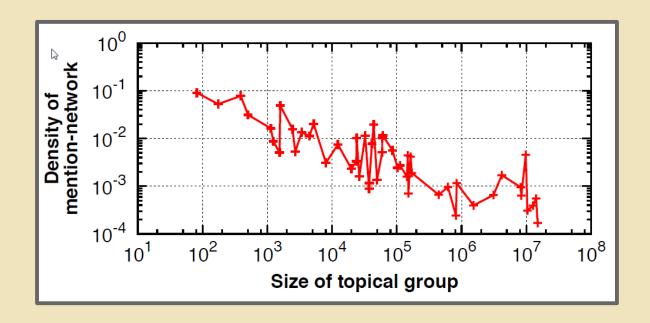
For niche topics, expert urls are 10 times more on topic.

Expert Proximity



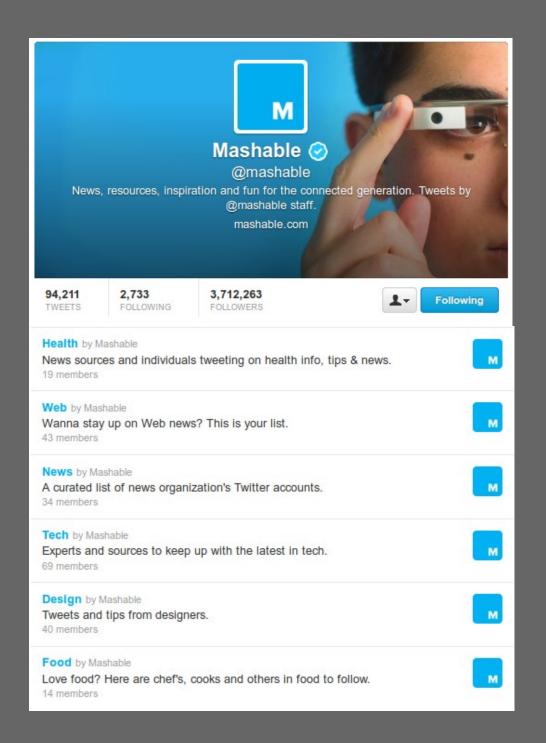
Experts are within two hops of 60-80% other experts.

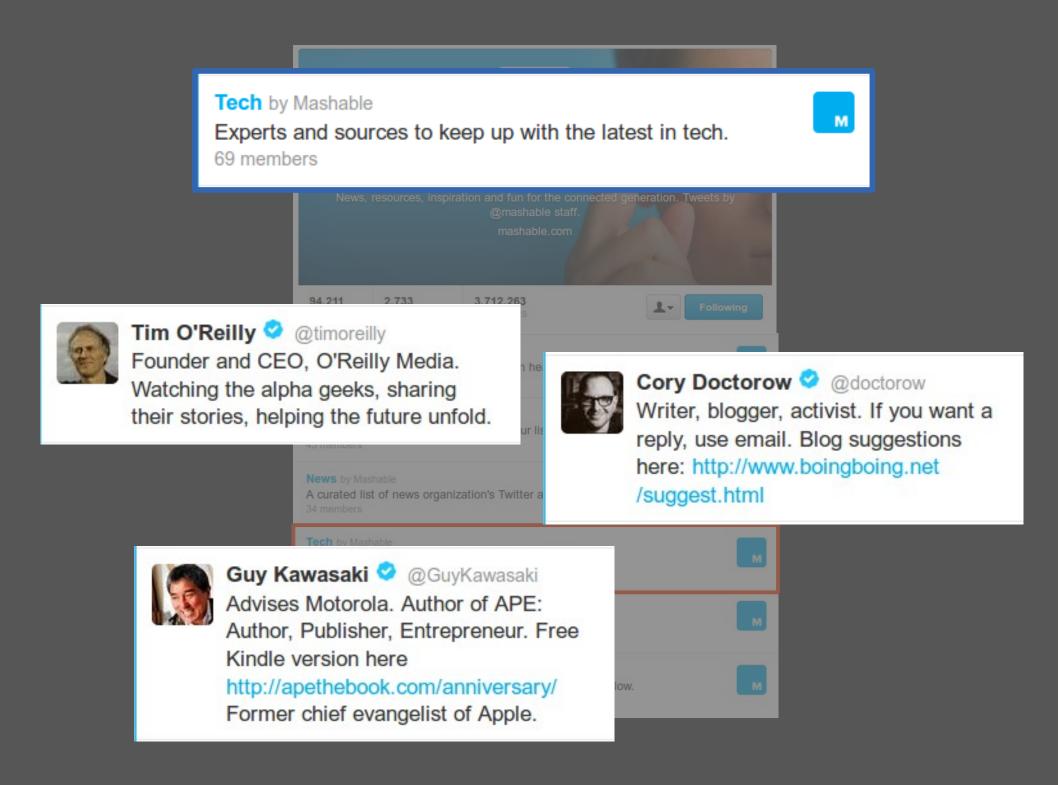
Density of Expert Mention Network

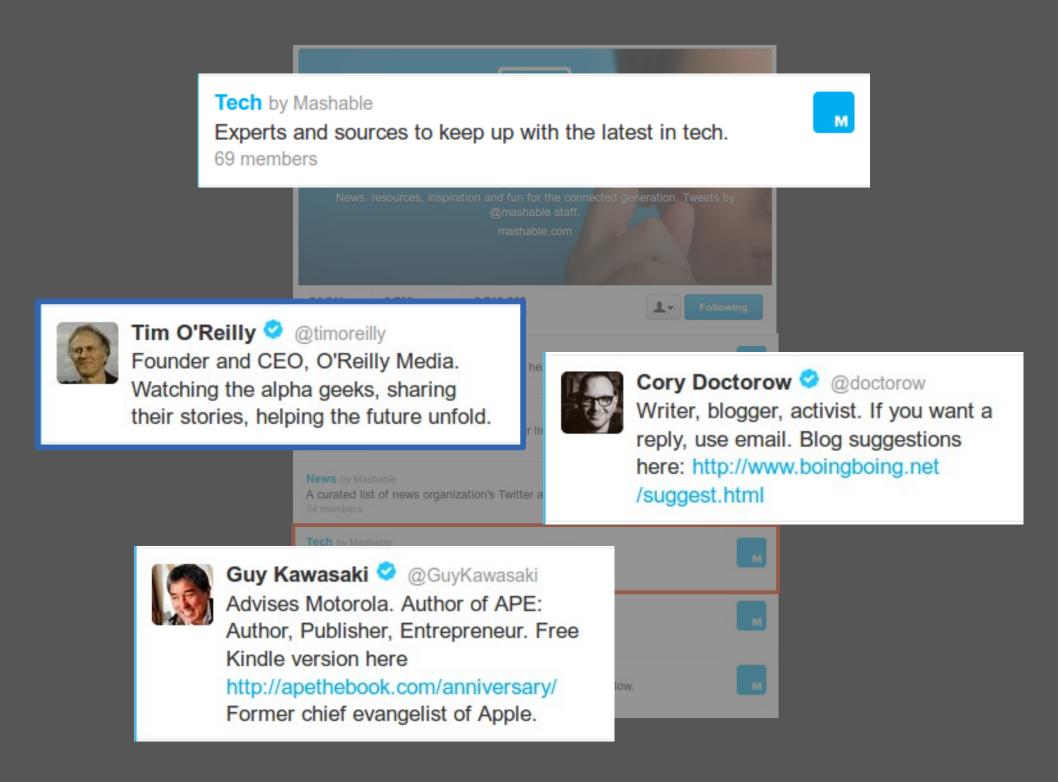


Destiny of mentions is much lower than connections.

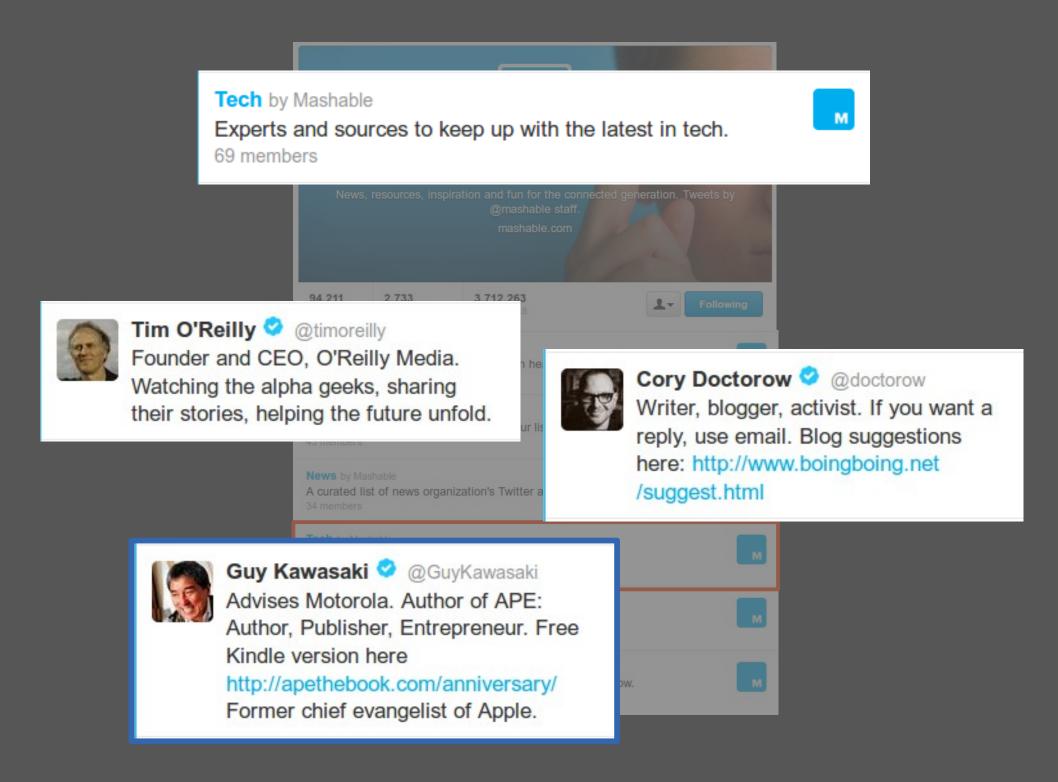
Mashable Lists

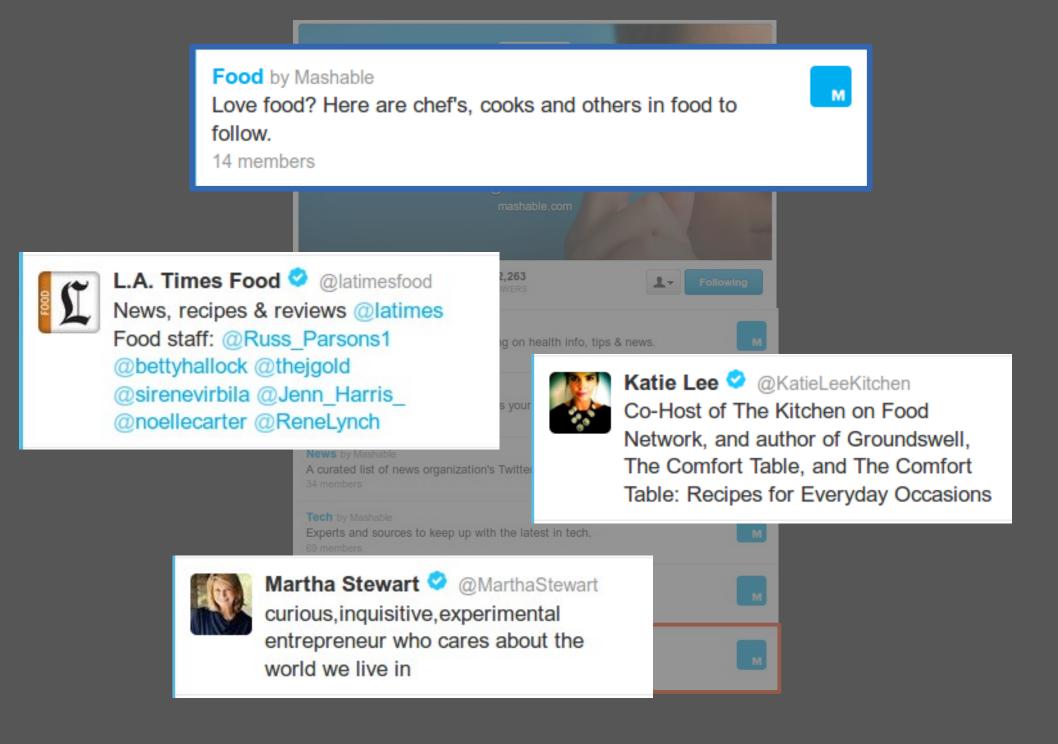


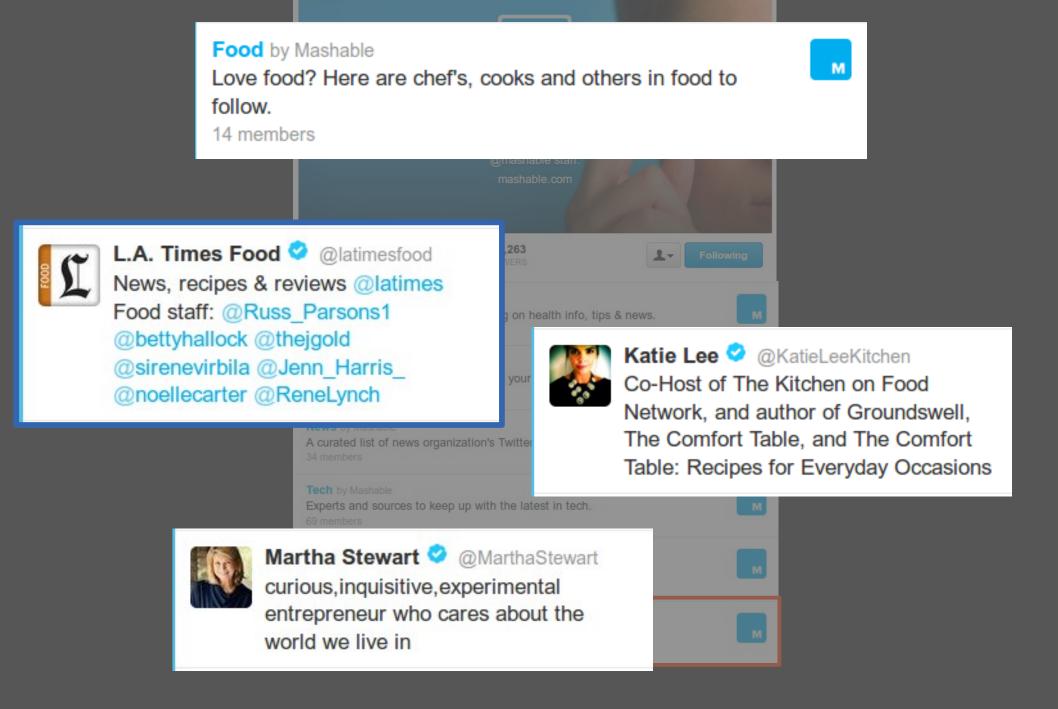


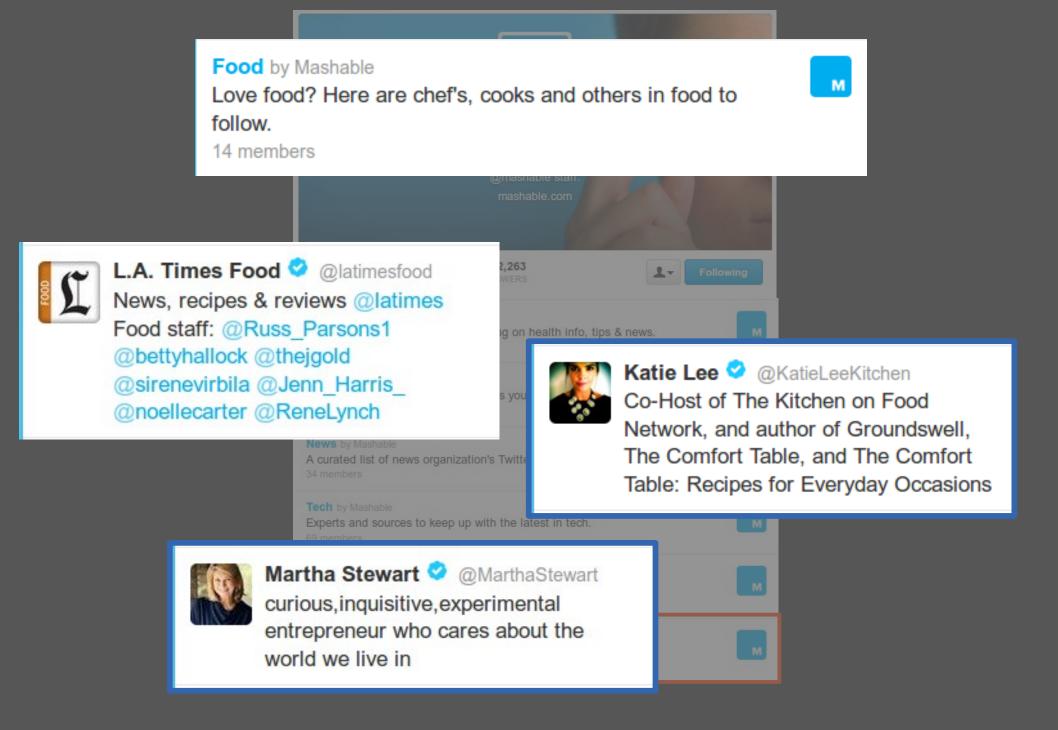














BarackObama : Barack Obama 📀

This account is run by Organizing for Action staff. Tweets from the President are signed -bo.

media vip president famosos leaders noticias america foreign tv

entertainment government personalidades inspiration international

news famous celebs

usa information obama figures important info barack stars business english

world politics

"Cognos: Crowdsourcing Search for Topic Experts in Microblogs"

Ghosh et. al, SIGIR, 2012.