



Understanding Social Media

Ricardo Baeza-Yates

Alejandro Jaimes

Yahoo! Research, Barcelona

- Information value is not
 - Static
 - Absolute
 - (always) Individual

- Relationships are not
 - Discrete
 - Fixed
 - Explicit

- Interactions, transfer, and effect vary in
 - Meaning
 - Content
 - Scope

Main Research Questions

- Information value: beyond information theory
- Social structure: culture and context dependent
- Interaction, transfer, and effect: much deeper analysis of human-human communications

Big Data: Some Yahoo! Numbers

»» **600+ Million Users**

»» **368 Million** PEOPLE VISIT THE YAHOO! HOMEPAGE MONTHLY

»» **9 Billion** ADVERTISEMENTS SERVED EACH DAY

»» **4.5 Billion** PAGE VIEWS DAILY

»» **#1** IN PAGE VIEWS AND MAIL

»» **100 Billion** MESSAGES SENT FROM 300 MIL. MAIL USERS MONTHLY

»» **#1** IN SPORTS, NEWS, FINANCE, AND ENTERTAINMENT

»» **81 Billion** MESSAGES SENT FROM 112 MILLION Y! MESSENGER USERS

»» **95 Million** FLICKR PHOTOS UPLOADED EACH MONTH

»» **70 Billion** MINUTES SPENT MONTHLY ON COMMUNICATIONS PROPERTIES



Further Motivation

- Why a social media site becomes successful?
- Can we answer this question if we do not know all the failures?
- Understanding implies:
 - Characterizing the process: information & relationships
 - Analyzing the dynamics: interaction
 - Predicting events: modeling
- Current research at Barcelona
 - Understanding Flickr & Twitter
 - Information spreading models

Placing Flickr Photos on a Map



[Home](#) [The Tour](#) [Sign Up](#) [Explore](#) [Upload](#)

You aren't signed in [Sign In](#) [Help](#)

[Search](#)

[Favorite](#) [Actions](#) [Share this](#)

[← Newer](#) [Older →](#)



Columnas UAB

Presentada a DPChallenge

Comments and faves



Bellprat (32 months ago)
Sembla que ballin



By **pasaro**
Pablo Saludes Rodil

This photo was taken on January 27, 2008 in Universitat Autònoma de Barcelona, Catalonia, ES, using a Canon PowerShot A710 IS.



1111 129 views 2 comments 1 favorite

This photo belongs to

[pasaro's photostream](#) (131)



This photo also appears in

- [Barcelona](#) (set)
- [Catalunya, next state in Europe](#) (group)
- [Barcelona](#) (group)
- [Barcelona Pics](#) (group)
- [1-5 favourites](#) (group)
- [Travel Photography](#) (group)
- ...and 5 more groups

Tags



Placing Flickr Photos on a Map

flickr® from YAHOO!

Home The Tour Sign Up Explore Upload

You aren't signed in Sign In Help

Search

Favorite Actions Share this

Newer Older



UAB - Ciencias Politicas

donde estoy estudiando!

puto ferrocarrilll

;-)

Comments and faves



Dogajolo - Lù pro (7 months ago)

Bravo ragazzo,

bravo ragazzo.

w.

By ..utopiacere..
Michele Lapini

This photo was taken on April 30, 2010 using a Nikon D80.

37 views 1 comment

This photo belongs to

..utopiacere..'s photostream

This photo also appears in

UAB (group: 225)



▶ España (group: 144,721)

▶ Catalunya, next state in Europe (group)

▶ Barcelona Pics (group)

▶ Expat in Barcelona (group)

▶ artistic photos (group: 65,961)

...and 5 more groups

Tags

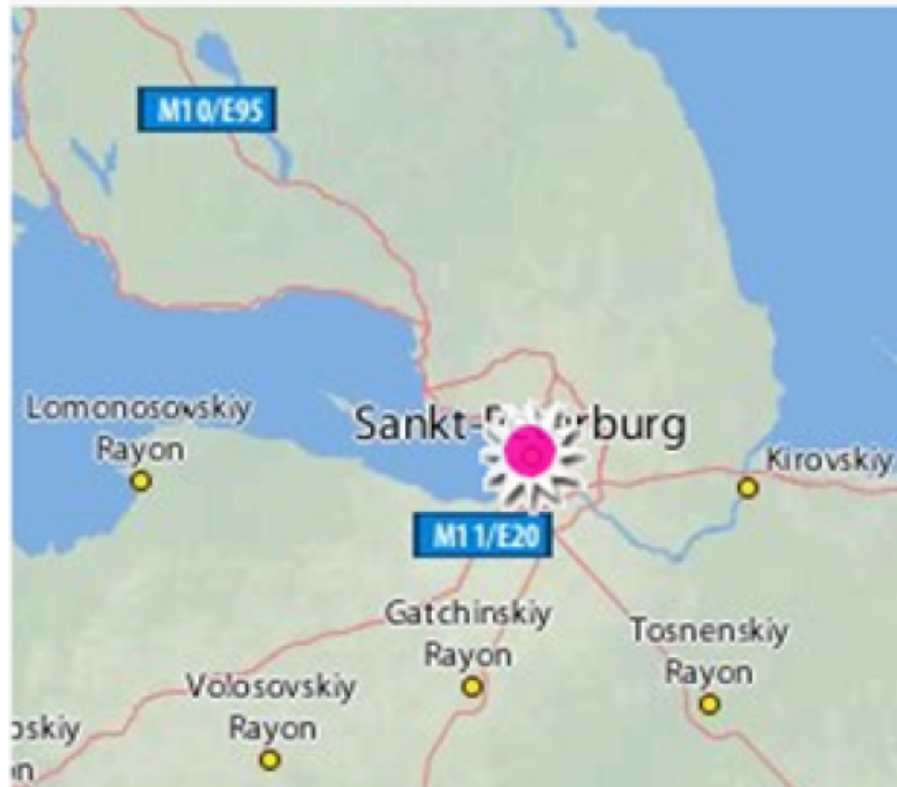
nikon • lapo • utopiacere • barcelona • uab •
universidad autonoma barcelona • università •
autonoma • universitat • universidad •
ciencias • politicas • sociologia • scienze •
politiche • campus • architettura • palazzo •
building • bellaterra • cerdanyola • sabadell •
ferrocarril • 10mm

But not a trivial problem!



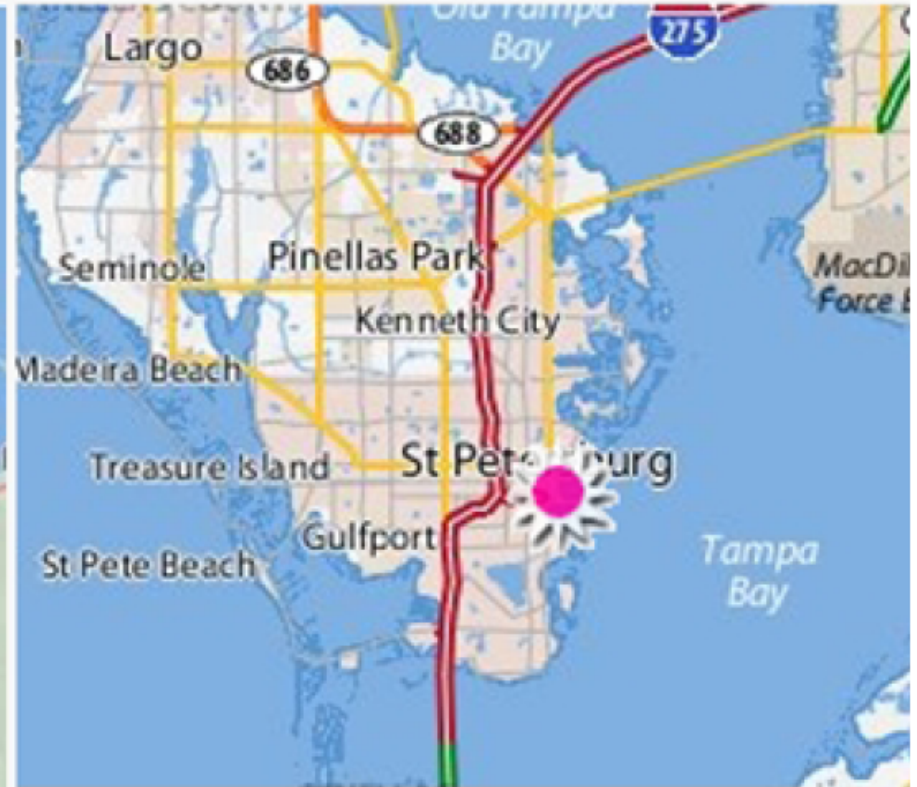
Placing Flickr Photos on a Map

St. Petersburg



All time popular tags
russia, church, bridge, cathedral, street, clouds, river, light, neva, saintpetersburg, petersburg, water, city, square, night, building, reflection, hermitage, winter, russian

St. Petersburg



All time popular tags
florida, reflection, pier, sunrise, stpete, tampa, st, tampabay, vinoypark, park, pelican, water, warpedtour, tattoo, bird, petersburg, marina, bay, egret, vinoy



Placing Flickr Photos on a Map

$$P(L|T)$$



Want to predict the location ... given the tag set

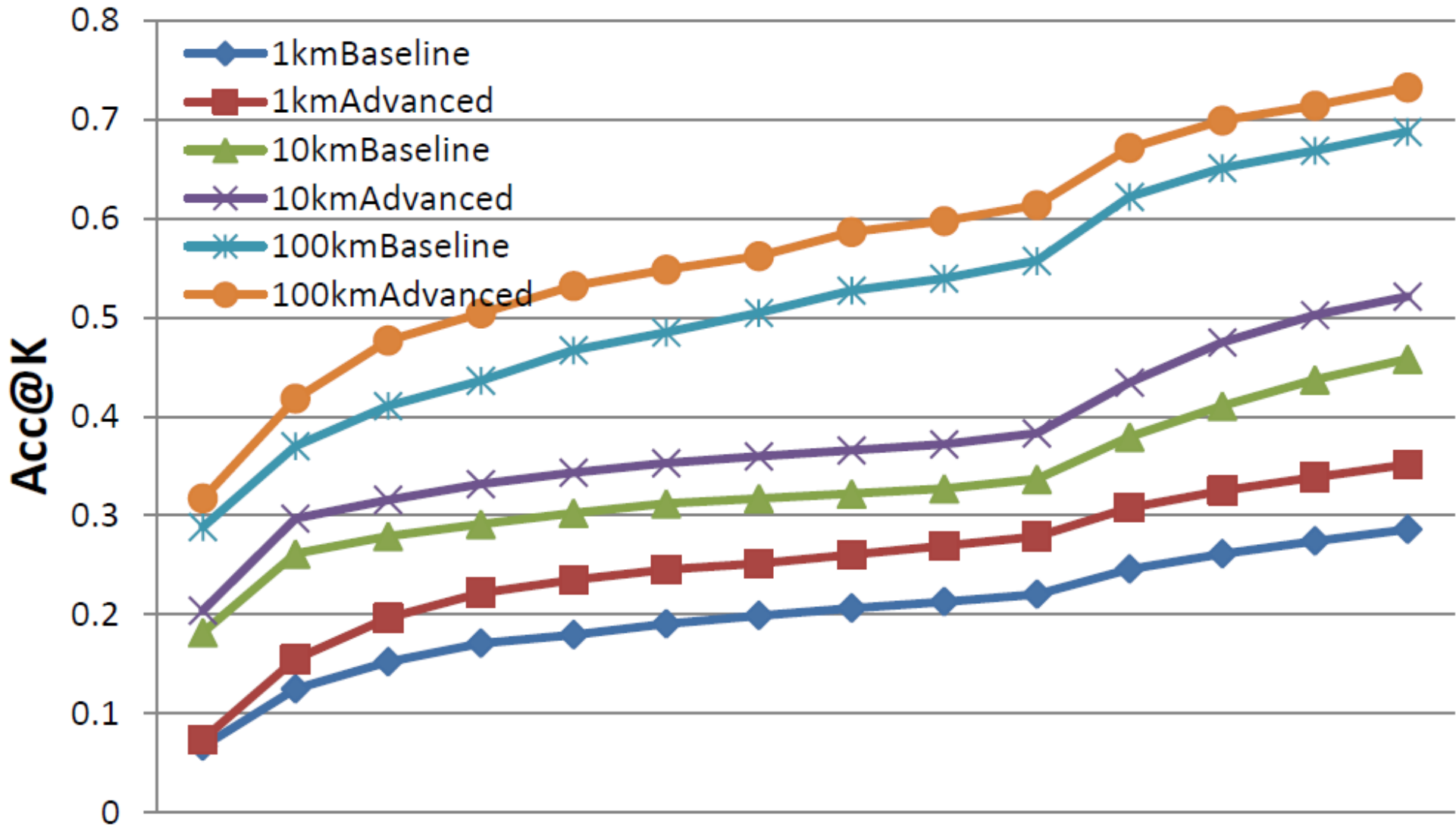
$$P(T|L) = \prod_{i=1}^{|T|} P(t_i|L)$$

Sparsity!

“smoothing”

Work by Vanessa Murdock and his team

Placing Flickr Photos on a Map



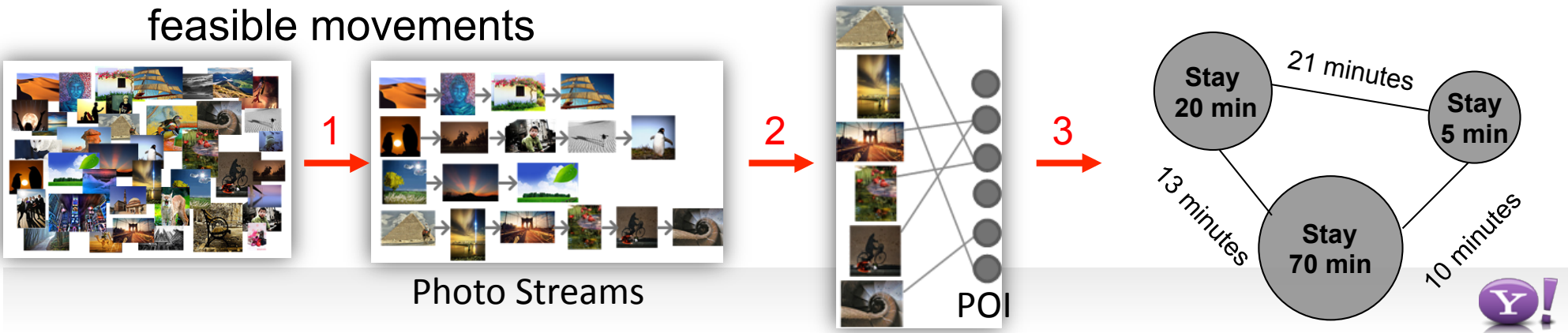
Automatic Construction of Travel Itineraries using Social Breadcrumbs

- **Goal:** construct intra-city travel itineraries automatically by tapping a latent source reflecting geo-temporal breadcrumbs left by tourists: Flickr

- **Approach**

- Extract photo streams of individual users (1)
- Associate points of interest with the photos (2)
- Using the timestamps of the photos, determine:
 - Lower bounds on the amount of time spent at a POI
 - Upper bounds on the transit time between POIs
- aggregate all user photo streams into a POI graph (3); apply the Orienteering algorithm to construct itineraries reflecting popular and feasible movements

Favorable results
as per
MT User Study,
comparing with
“ground truth”
itineraries



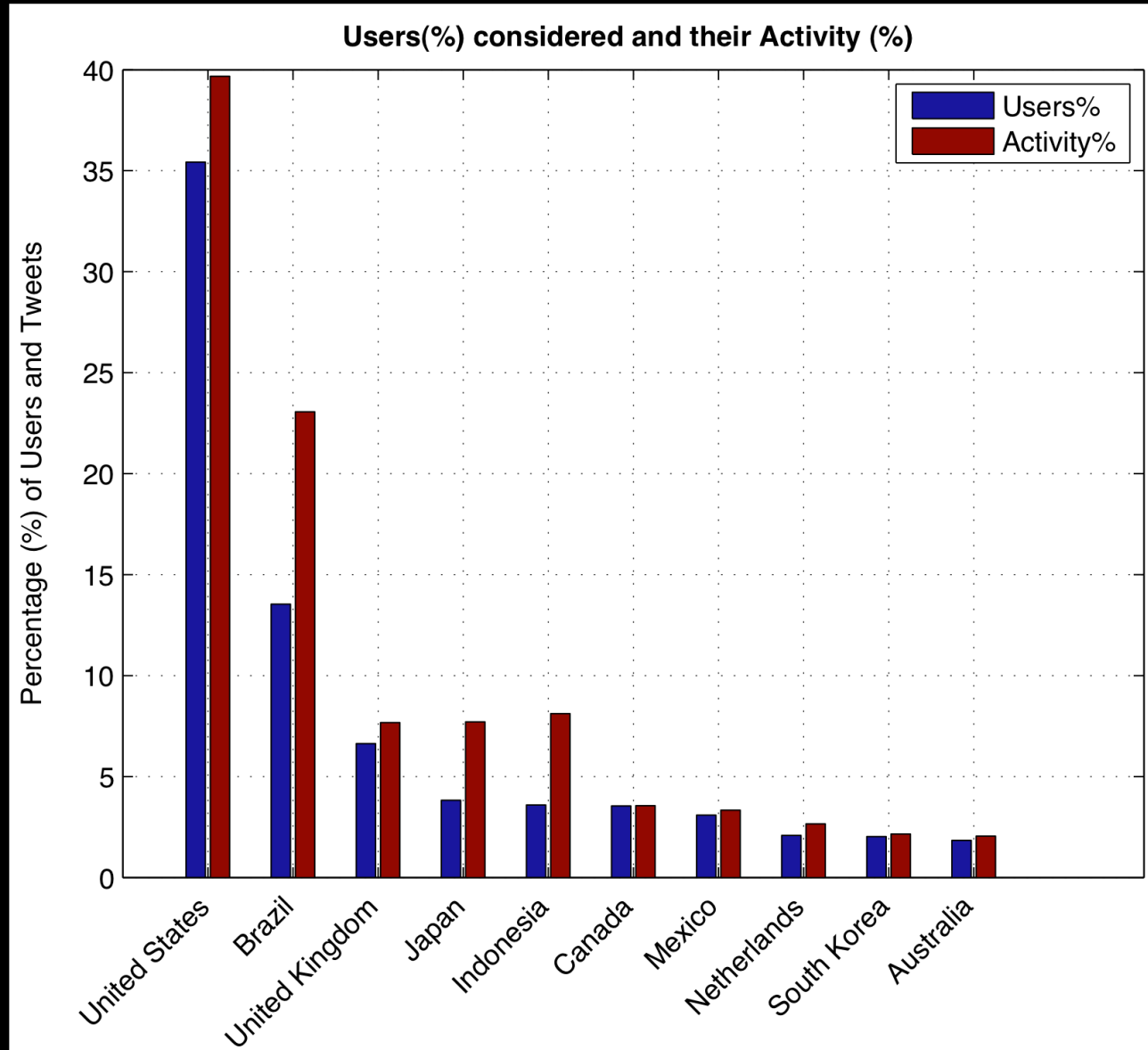
Microblogging Behavior

[A. Jaimes, R. Garcia, B. Poblete, M. Mendoza, ACM WebSci '11]

- 550 M tweets examined over a period of 9 days (76 M users)
 - Location information extracted
- All tweets for one year, for 10 most active countries
 - 4.7 M users, 5.2 Billion tweets

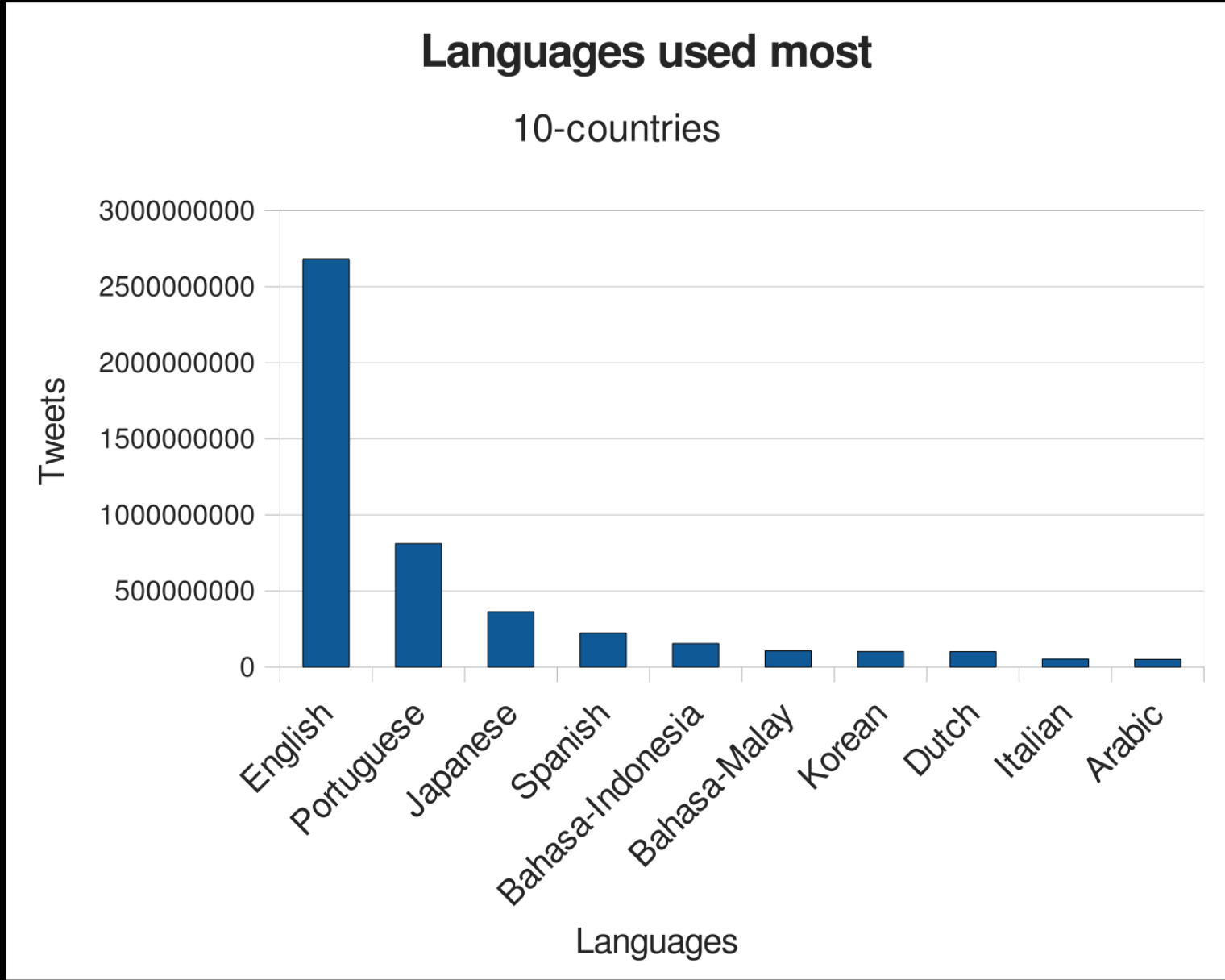
Microblogging Behavior

Results I



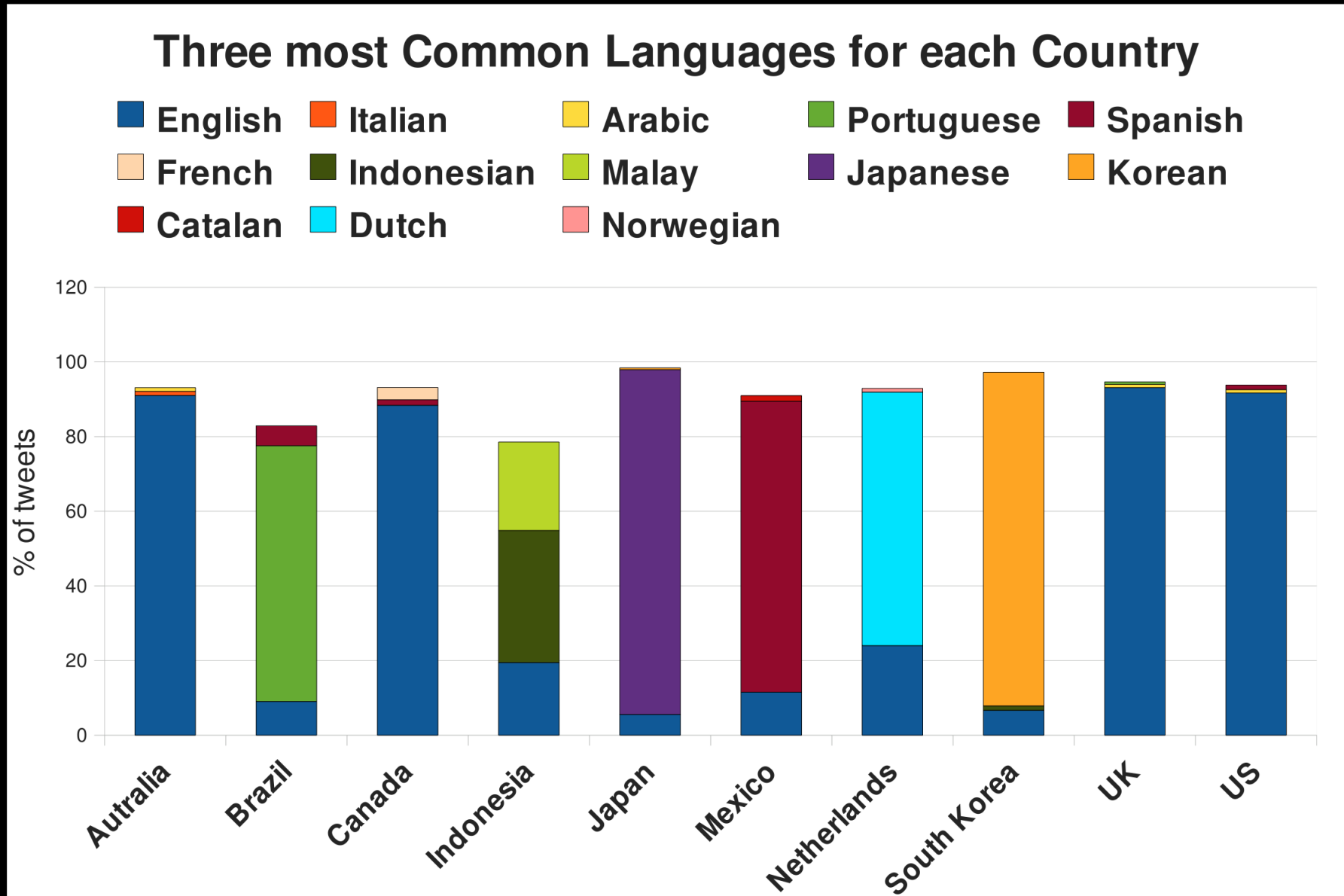
Microblogging Behavior

Results II



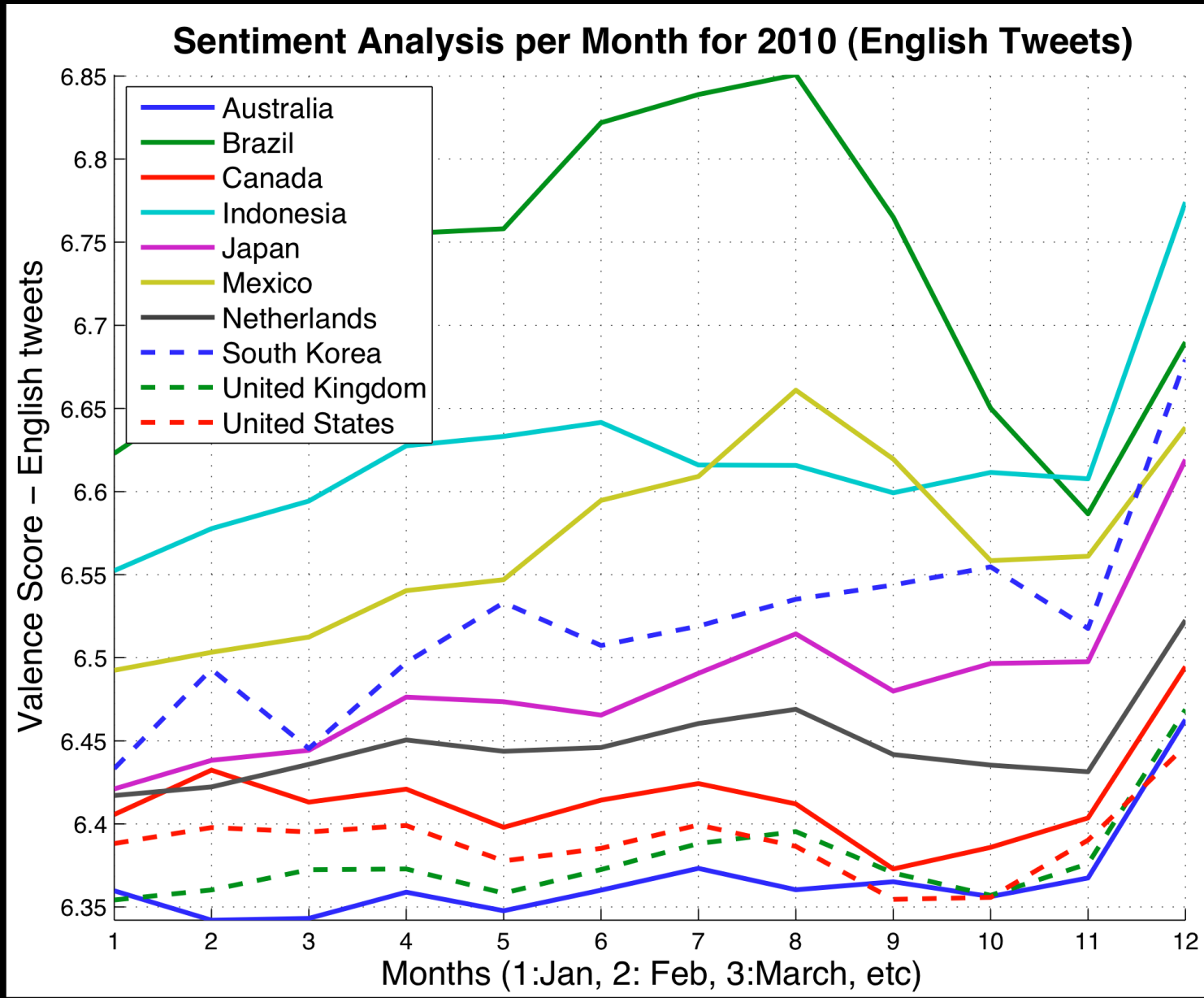
Microblogging Behavior

Results III



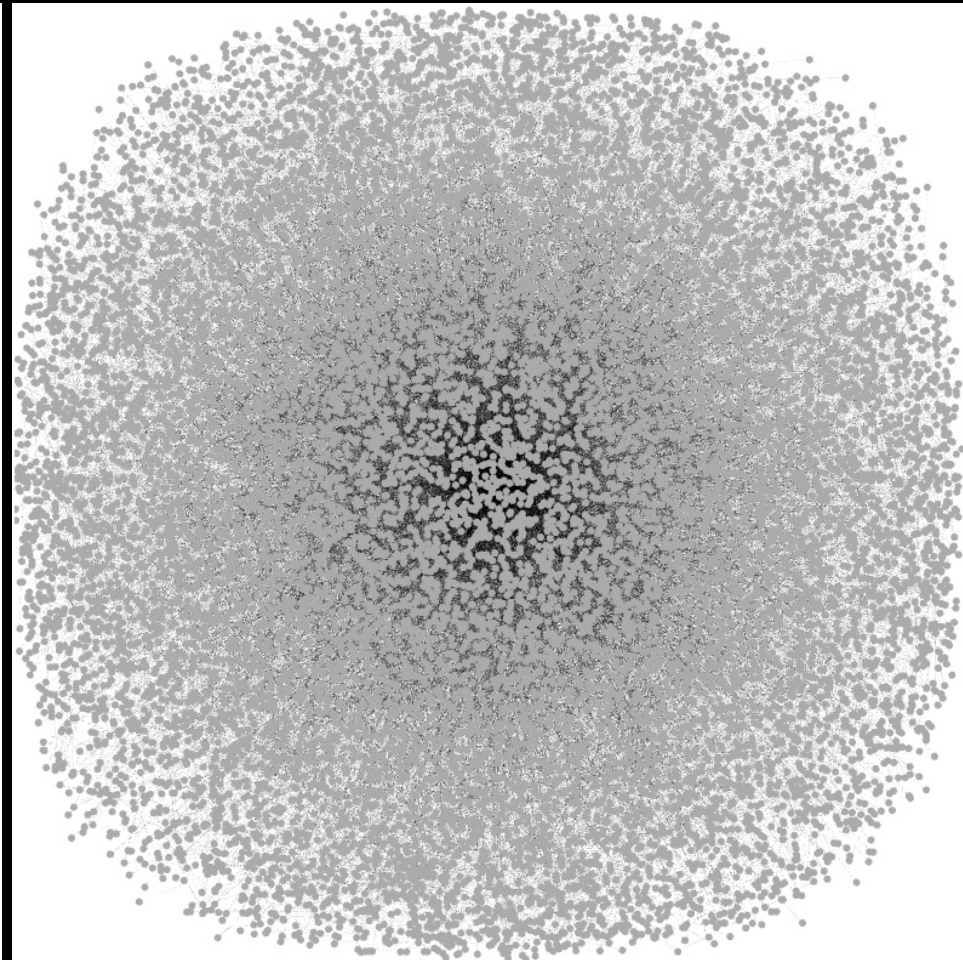
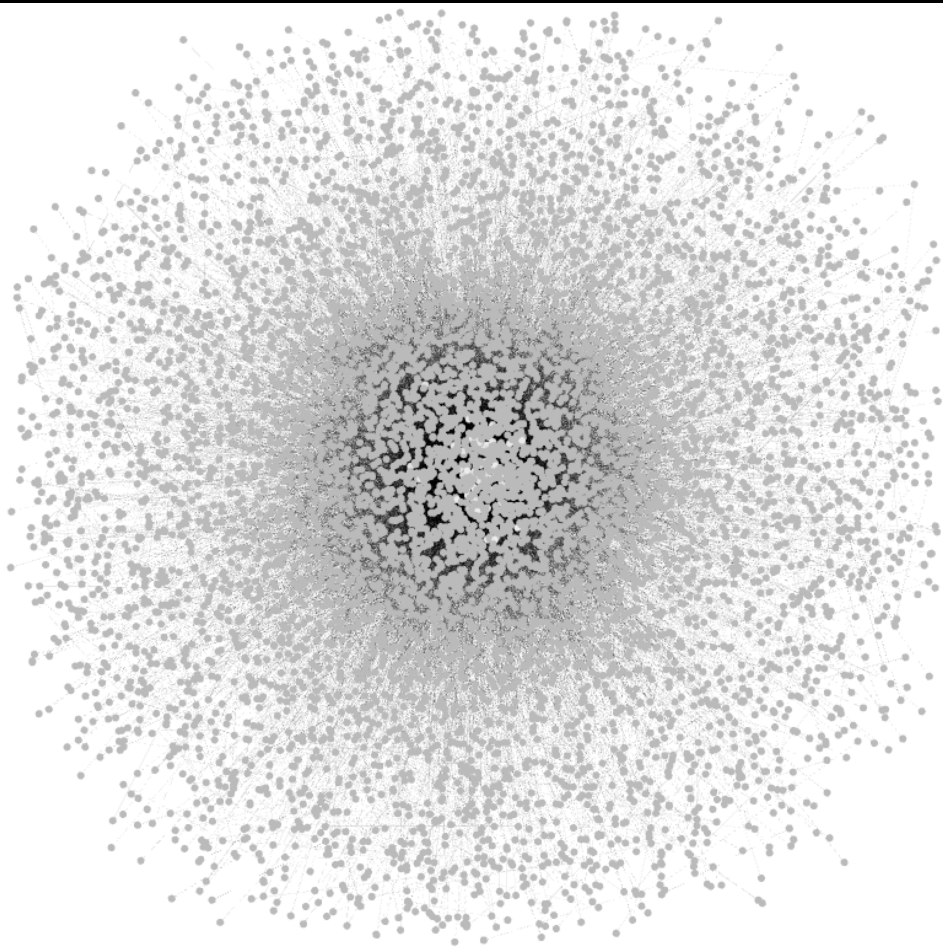
Microblogging Behavior

Results IV



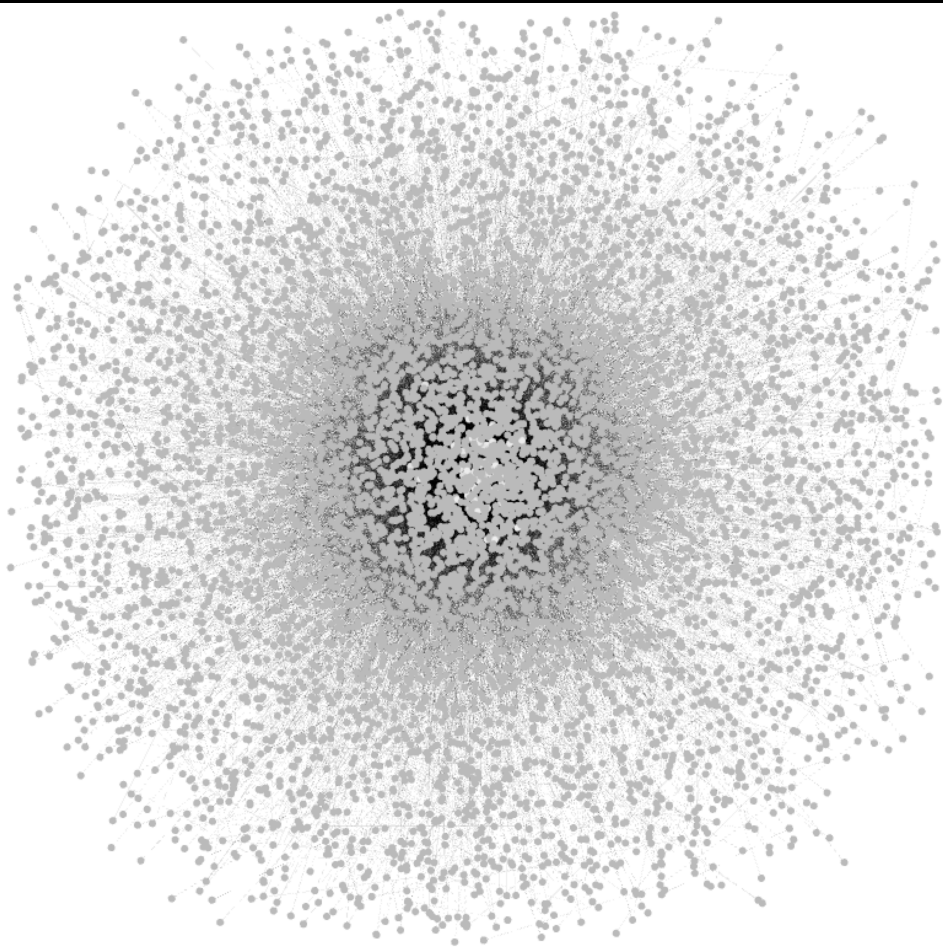
Microblogging Behavior

**Can you guess which is Australia
and which is Indonesia?**

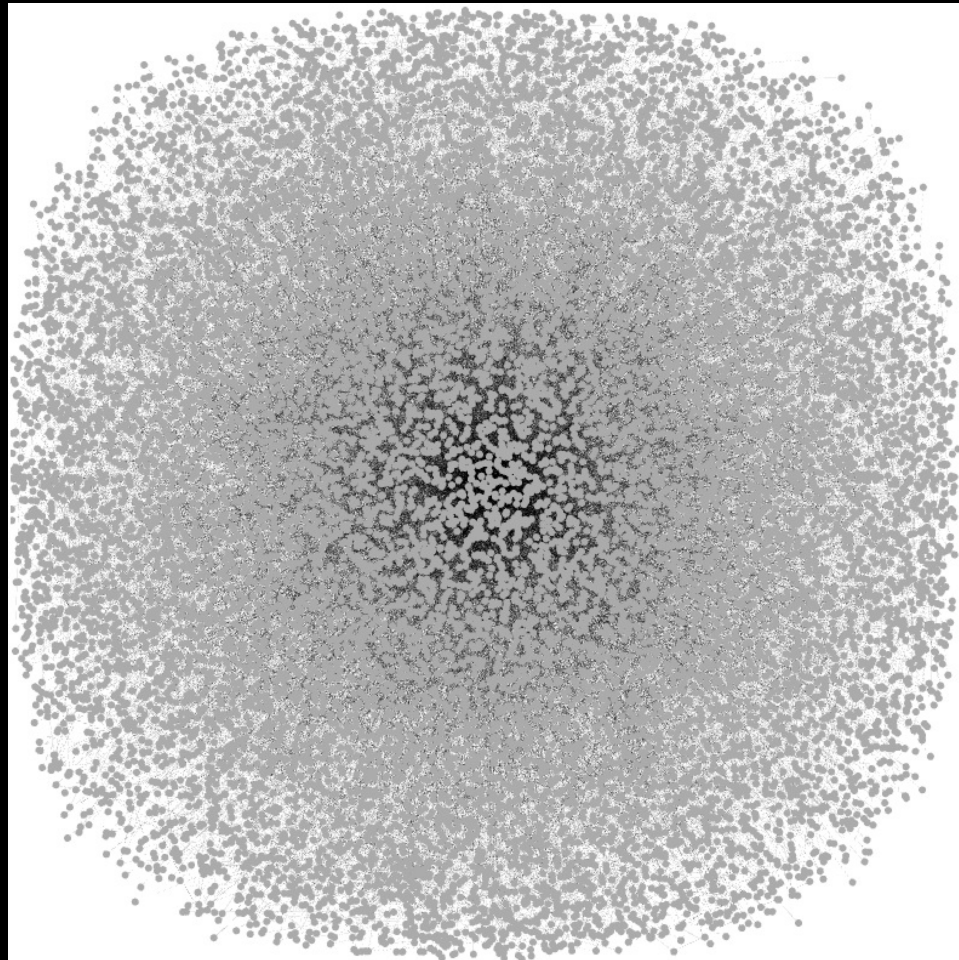


Microblogging Behavior

Australia

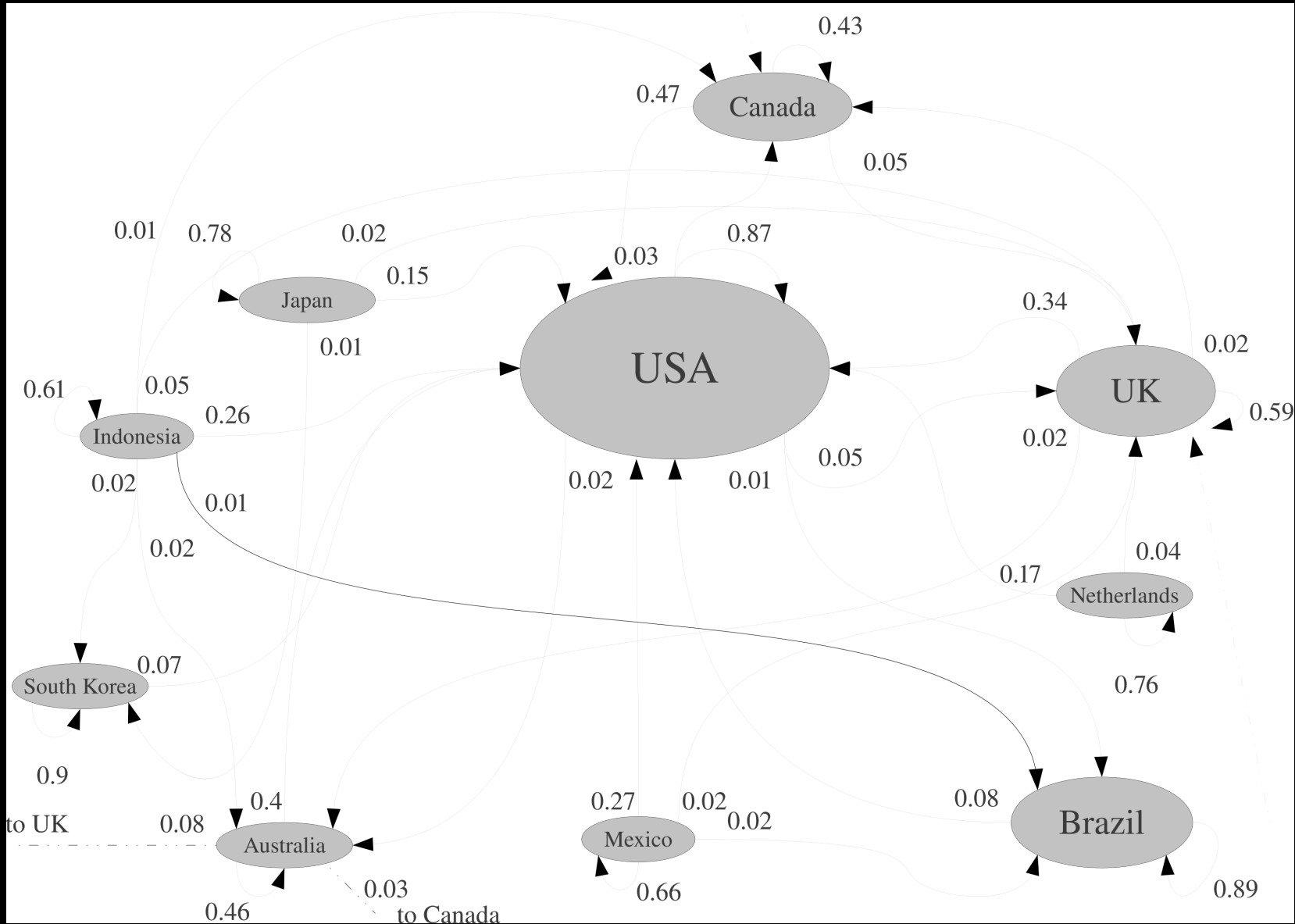


Indonesia



Microblogging Behavior

Ties Within Countries



Microblogging Behavior

Reciprocity

Country	Users	Cov.(%)	Links	Cov.(%)	Reciprocity(%)
USA	1,616,702	12.47	11,310,538	12.46	18.91
Brazil	688,427	5.31	4,248,259	4.68	13.49
UK	286,520	2.21	1,370,699	1.51	17.22
Japan	133,536	1.03	408,486	0.45	32.01
Canada	132,240	1.02	553,726	0.61	26.11
Indonesia	130,943	1.01	199,704	0.22	26.97
Mexico	112,793	0.87	399,409	0.44	17.27
Netherlands	86,863	0.67	354,021	0.39	22.11
South Korea	80,381	0.62	499,261	0.55	28.14
Australia	67,416	0.52	299,556	0.33	23.51

Twitter under Crisis

twitter

Home Profile Find People Settings Help Sign out

What's happening?

140

There's looting at the supermarket in X.

jesseluna Latest: Video - Water receding 10 ft in Peru <http://bit.ly/cFT505>
(h/t @TM_Video @Magiko) #sunami #terremoto Retweeted by you 21 minutes ago

update

True? False?

terremoto-chile

Following: 21 | Followers: 0

View list page >



adnradio Avión de equipo israelí se desvió a Perú por el terremoto
<http://url4.eu/1Ubc8>

less than 20 seconds ago via RSS2Twitter



adnradio ANFP decidió suspender toda la jornada del fin de semana debido al terremoto <http://url4.eu/1Ubc9>

less than 20 seconds ago via RSS2Twitter



adnradio Falta de locomoción afecta a la capital y Metro suspendió servicio <http://url4.eu/1Ubc7>

less than 20 seconds ago via RSS2Twitter



TVN24H Comisión europea destinará 3 millones de euros para ayudar a Chile. Concepción sin luz y agua. #24h #terremotochile

half a minute ago via web



CruzRojainforma Tomar medidas por posibles apagones en la noche, a mano velas, fósforos, linternas, radio m, recolecte agua

2 minutes ago via mobile web



kikegiles

3,247 tweets

1,496 following 1,367 followers 117 listed

NBC·Twitter·Tracker

n. What the world is saying about the Olympic Games.

Home

@kikegiles

Direct Messages 557

Favorites

Retweets

Search

Saved Searches

#ebe09

#w2e

#frasedeldia

Lists

terremoto-chile >

prensa-peru

educacion

prensa-extranjeros

famosos

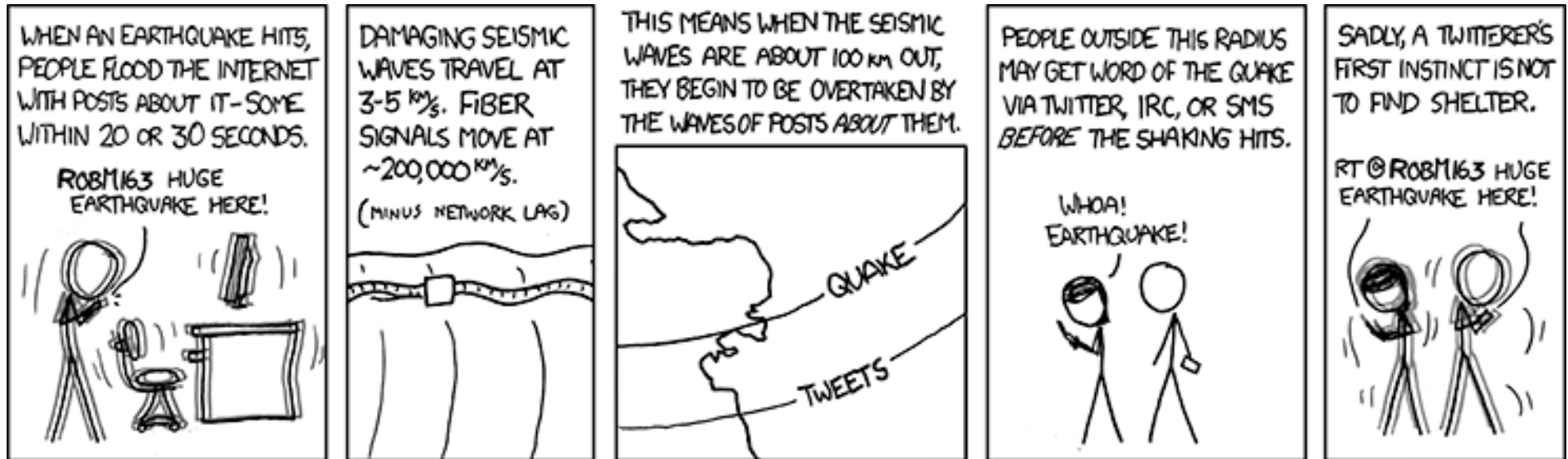


Information Credibility in Twitter

Carlos Castillo, Marcelo Mendoza, Bárbara Poblete

Yahoo! Research, Barcelona & Santiago de Chile

<http://xkcd.com/723/>



Twitter helped, but ...

- Large majority of tweets were very helpful
- Some tweets were not
 - False tsunami warnings
 - False reports of looting
 - ...

A “small” earthquake in Chile

- Sat Feb 27, 2010. 03:34 local time
- 7th largest recorded in history
 - 8.8 Mw, 90 s, 500-600 casualties
 - Haiti: 7.7 Mw, 60 s, 90k-300k casualties
- Communications
 - Almost impossible for 2-3 hours
 - First video images 6-7 hours after quake



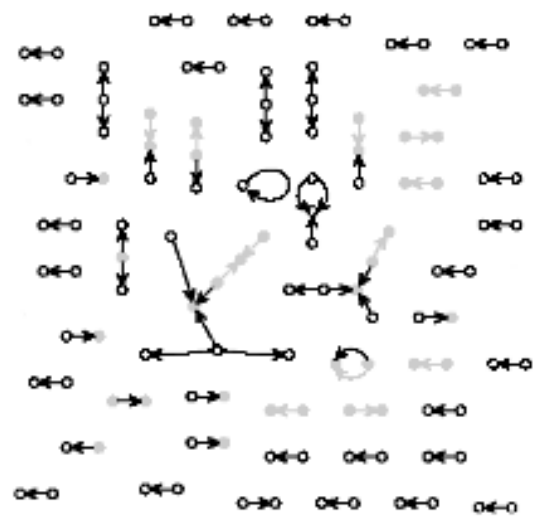


Chileans love Twitter

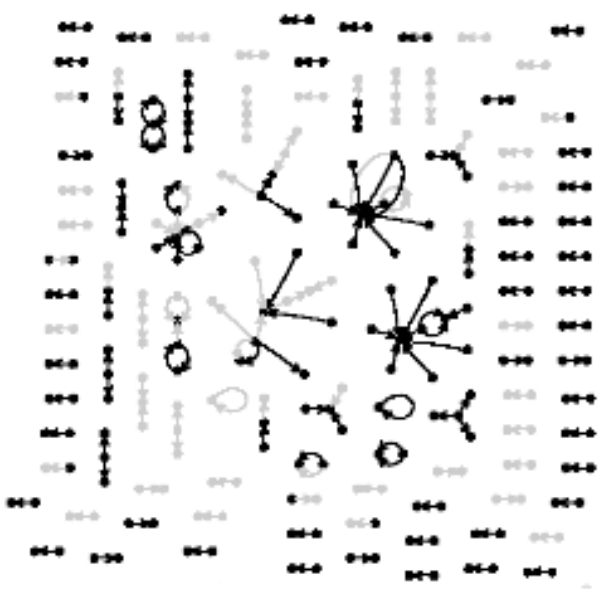
- Prominent role for communications
 - online and offline
- All public figures tweet
- Well integrated with traditional media
 - Since ~1yr before earthquake (elections)



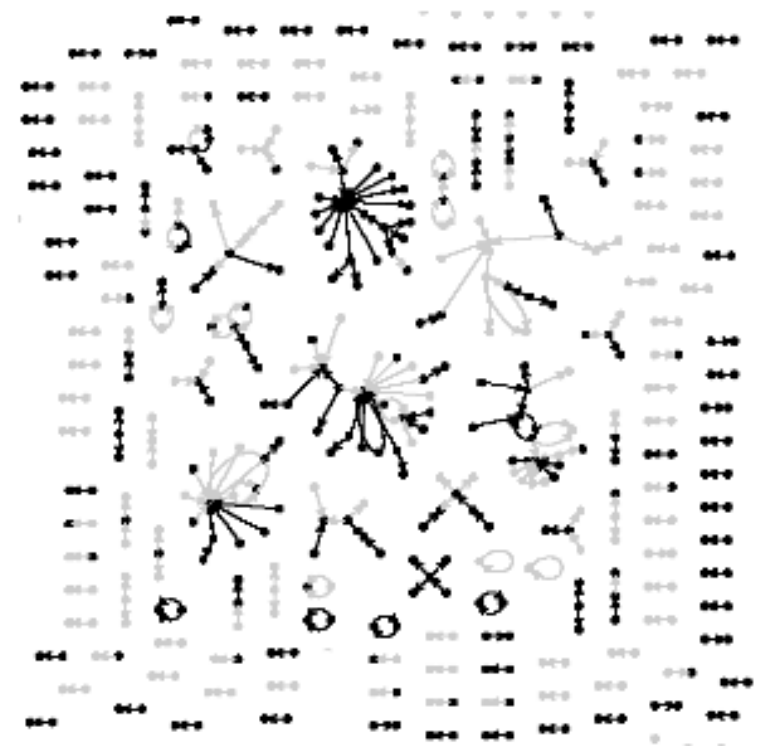
(a) 03:35:00 - 03:49:00



(b) 03:50:00 - 04:04:00



(c) 04:05:00 - 04:19:00



(d) 04:20:00 - 04:34:00



(a) 27 Feb



(b) 28 Feb



(c) 01 Mar



(d) 02 Mar

Manual classification results

Case	# of unique tweets	% of re-tweets	# of unique "affirms"	# of unique "denies"	# of unique "questions"
Confirmed true TRUE					
The international airport in Santiago is closed	201	41	291	0	7
The <i>Viña del Mar International Song Festival</i>	256		256	0	3
Fire in the Chemistry Faculty at the University of Concepcion	42	49	38	0	4
Navy acknowledges mistake informing about tsunami warning	107	30	107	4	6
Small aircraft with six people crashes near Concepción				0	4
Looting of supermarket in Concepción	160	44	149	0	2
Tsunami in Iloca and Duao towns	153	32	140	0	4
TOTAL	1181		1123	4	30
AVERAGE	168,71		160,43	0,57	4,29
False rumors FALSE					
Death of artist Luciano Nájera	50	37	24	12	8
Tsunami warning in Valparaiso	700	4	45	605	27
Large water tower broken in Rancagua	126	43	62	38	20
Cousin of football player Gary Medel is a victim	94	4	44	34	2
Looting in some districts in Santiago	250	37	218	2	20
"Huascar" vessel missing in Talcahuano	234	36	54	66	63
Villarrica volcano has become active	228	21	55	79	76
TOTAL	1682		502	836	216
AVERAGE	240,29		71,71	119,43	30,86

How many people confirmed?

How many people denied?

Gun alert on oxford circus right now true?

11:25am Jan 19th 2011 via TweetDeck



[HannahLisaGreen](#)
Hannah Green

has been told to stay in the office as the shooting in Oxford Circus... the cirens gone past so is it true...?

11:30am Jan 19th 2011 via web



[katie_raby](#)
Katie Raby

The penny is starting to drop about Oxford Circus. This has been like shooting fish in a barrel.

11:44am Jan 19th 2011 via web



[smorgasbord](#)
smorgasbord

Gunman loose in Oxford Circus?
[@StephAldrich](#)

11:26am Jan 19th 2011 via web



[kellymorlock](#)
Kelly Morlock

There's some kind of shooting going on in Oxford Circus...some rumour about a gunman....trying to find out more details

11:30am Jan 19th 2011 via TweetDeck



[gromitski](#)
Stu Collett

PLEASE RETWEET - Police say incident on Oxford Circus/ St is a rumour started by someone.... #Idiot

11:46am Jan 19th 2011 via web



[jamie_shepherd](#)
Jamie Shepherd

is there a shooting in Oxford Circus now???? this is what i'm hearing

11:26am Jan 19th 2011 via TweetDeck



[gpcrc](#)
Grant R-C

Anyone know if there really is something going on in Oxford Circus, or if it's a hoax/incorrect rumour?

11:31am Jan 19th 2011 via TweetDeck



[melstrutt](#)
Melanie Strutt

Twitter: a social networking & microblogging service that helps breed and incite panic with conflicting reports of a gunman at Oxford Circus

11:46am Jan 19th 2011 via web



[RealSamRobson](#)
Sam Robson

shooting in oxford circus - talk of the True or Hoax?

11:27am Jan 19th 2011 via web



[Pheebie](#)
Phoebe Campbell

There's been a shooting at Oxford Circus

11:31am Jan 19th 2011 via web



[racheljanejones](#)
Rachel Jones

I've just been to oxford circus where this shooting is supposed to have been, not 30 seconds ago, and it all looks perfectly normal..

11:46am Jan 19th 2011 via web



[Roslovesme](#)
Alice Carlisle

**Can we automatically
detect false tweets?**



Can we automatically
detect false tweets?

Of course not.



Information credibility

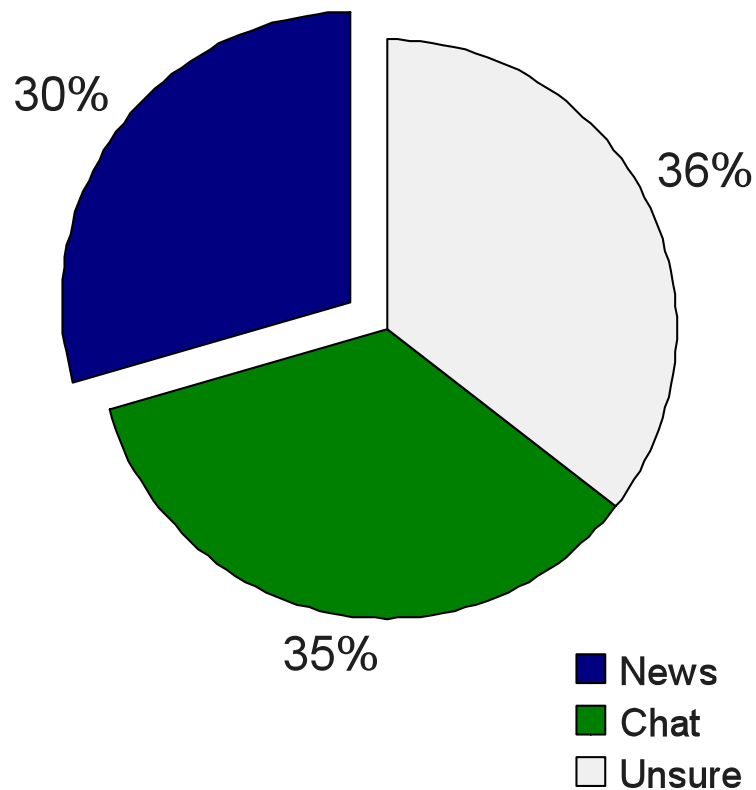
- Perceived quality
- Made of multiple dimensions
- Examples:
 - Same article written by male/female author
 - Same headline in news/twitter
 - ...

Supervised classification

- Events (tweet sets) from TwitterMonitor
 - [Mathioudakis & Koudas 2010]
- Labels from Amazon's Mechanical Turk
- Decision trees

Labeling: News or Chat

- 383 events from TwitterMonitor.net [Mathioudakis & Koudas]
- 7 evaluators per event
- ≥ 5 agreement



Manage HITS

Identifying news/events from tweets

Delete this HIT

Requester:	Marcelo Mendoza Rocha	Assignments Pending Review:	0
HIT Expiration Date:	Nov 30 2010, 06:11 AM PST	Reviewed Assignments:	0
Rewards:	\$0.06	Remaining Assignments:	7
Assignments Requested:	7	Remaining Time:	Expired Add time

Description: In this job, you will need to indicate if most of the tweets in a group are spreading the news about a specific EVENT/NEWS. You will be asked to summarize the topic behind the tweets in a short sentence.

Keywords: Twitter, event detection, news, summarization, research

Identifying specific news/events from a set of tweets

Guidelines:

Users of Twitter post short messages, each up to 140 characters, commonly known as tweets.

In this task you will need to indicate if most of the tweets in the group are:

1. Spreading news about a **specific news/event**
2. Comments or conversation

A **specific news/event** must meet the following requirements:

- be an affirmation about a fact or something that really happened.
- be of interest to others, not only for the friends of each user.

Tweets are not related to a **specific news/event** if they are:

- Purely based on personal/subjective opinions.
- Conversations/exchanges among friends.

- For each group, we provide a list of descriptive keywords that help you understand the topic behind the tweets.

Examples:

Specific news/event

- Study says social ad spending to reach \$1.68 billion this year
- Obama to sign \$600 million border security legislation <http://dlyt.it/3kqap>
- Huge brawl in GABP!!! #cardinals v #reds

Conversation/comments

- Probably should have brought rainboots to wort today. #regret
- Listening to @jaredleto performing Bad Romance gives me goosebumps
- Lovely weather for cats

Item 3.

Consider the following group of tweets:

- RT @jbreeze24 @blazetrilla lakrs bout to get raja bell &t;&t;&t;dat nigga a scrub anyway fuck dat nigga he gonns warm da bench up
- Fuck raja bell going to Utah? Damn!
- RT @jsharikavi: the #Utah #Mormons look like they are now getting raja bell.....&t; god u w fool
- SMH raja bell told Kobe Nevermind on meeting him and went to UTAH. dick move.
- @IRapedKOBEE raja bell definitely goin 2 da lakers. he'll b stupid not 2. #WeDaChamps
- @ChgTheGME they'll see what happens next year. Yo kinda mad raja bell went to the jazz instead of us
- Don't mind Shannon brown coming back would of preferred raja bell but brown works. I'm just happy farmer is gone and Lakers got @SteveBlakes
- @Basketball_Ron Ron what do you think about the lakers going after raja bell
- Fuck U raja bell ! U chose money over a chamionship w/ Kobe lol
- RT @Lockedonsports: O'Connor "we got someone who can guard the best perimeter defender and wants to" in raja bell

descriptive keywords: "raja", "bell"

The previous tweets are:

- spreading a specific news/event?
- conversation/comments among friends?

Please provide a description of the topic covered by the previous tweets in only one sentence:

[Empty text input box]

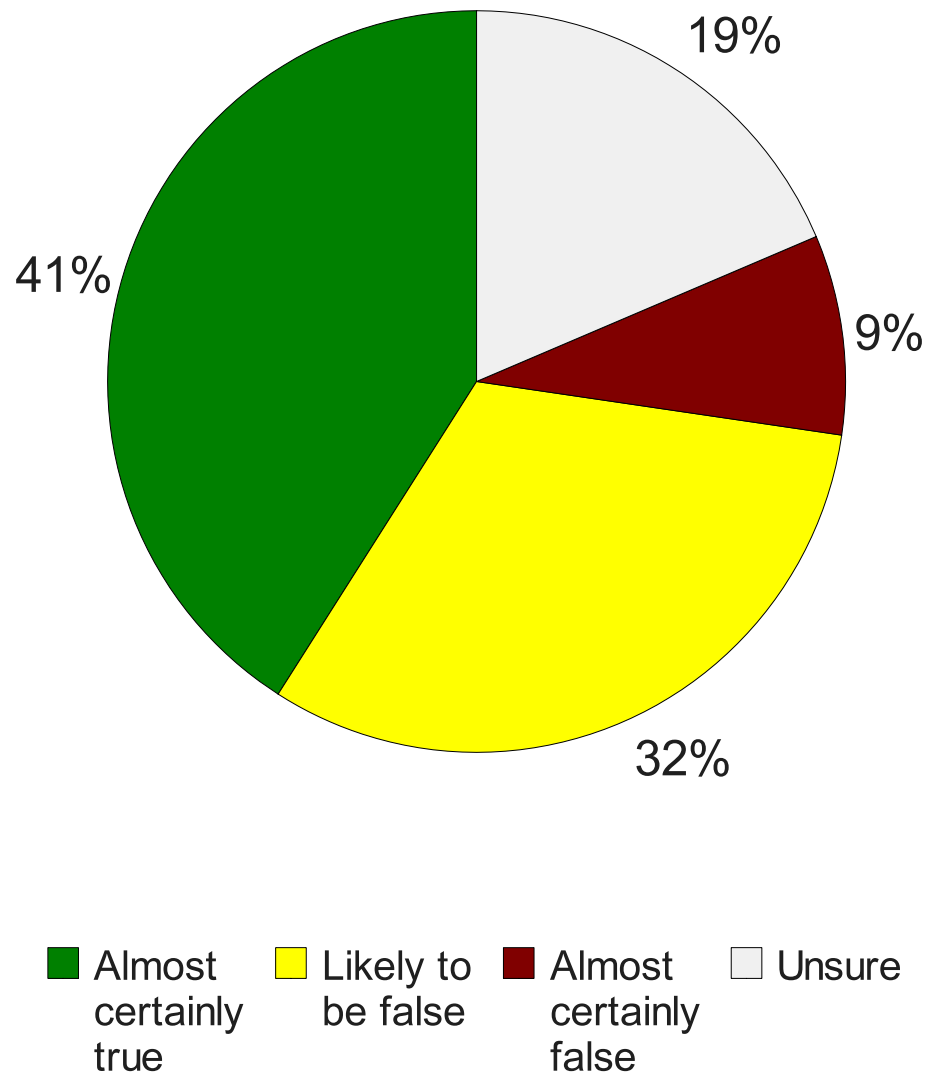
Spreading a specific news/event

-OR-

Conversation or comments among friends.

Labeling: Credible or Not

- 747 events automatically classified as news
- 7 evaluators per event
- ≥ 5 agreement



Manage HITs

Distinguishing credibility levels from a set of tweets

[Delete this HIT](#)

Requester:	Marcelo Mendoza Rocha	Assignments Pending Review:	0
HIT Expiration Date:	Dec 26 2010, 08:03 AM PST	Reviewed Assignments:	0
Reward:	\$0.05	Remaining Assignments:	10
Assignments Requested:	10	Remaining Time:	Expired Add time
Description:	In this task you will need to indicate credibility levels for the topic behind short messages in Twitter.		
Keywords:	Twitter, credibility, news, research, rumors		

Distinguishing credibility levels from a set of tweets

Guidelines:

Users of Twitter post short messages, each up to 140 characters, commonly known as tweets.

In this task you will need to indicate a level of credibility for the topic behind these short messages in Twitter

- We provide credibility levels: "almost certainly true", "likely to be false", "almost certainly false", and "I can't decide".
- For each group, we provide a short descriptive sentence that help you understand the topic behind the tweets. We provide also the date of the group of tweets.

Examples:

News

- \$1.20 trillion deficit for 2010 confirmed.
- Vimeo, an application, is now available on the iPad.
- Spain wins the 2010 FIFA world cup in extra time

Rumors

- Hurricane in the south of Chile
- Microsoft releases Office 2012
- Justin Bieber lyrics auctioned off for \$12 million

Item

Summary sentence: "underwood carrie"

Date: Sat Jul 10 2010

Sample of messages/tweets ordered by timeline:

- @istruckd_Annie I like all type of music from india arie , wale , kanye , carrie underwood. I like erbody :)Check this out: Carrie Underwood Wedding Takes Her Off The Market <http://www.notsorealnews.com/carrie-underwood-wedding/>
- [1. alexis cohen] [2. dorell wright] [3. carrie underwood wedding] [4. las tablas panama] [5. stephen colletti]
- congrats to my beautiful friend brittany and lovely hubby ryan on their wedding; oh and of course carrie underwood and mike fishers wedding!
- gonna need alot of \$ RT @sportschickblog carrie underwood married mike fisher today, not @shill910 ... i
- Carrie Underwood wedding {
- #np carrie underwood-temporary home
- bagoesss itu T.T hiksss RT @aayyyuuuu mandi aaahh..... #nowplaying I told you so - carrie underwood.... gak bosen2 aku dengerinnyaaaa
- Babs Says: Carrie Underwood and Mike Fisher Wed! <http://www.babblewood.com/2010/07/carrie-underwood-and-mike-fisher-wed/>
- carrie underwood got married...i have no reason to live...
- New pix from LAX of Carrie Underwood & Mike Fisher leaving for their honeymoon! <http://carrie-underwood.love.com/photos?photodeeplinkNum=0>

Please classify these messages as:

- Almost certainly true
- Likely to be false
- Almost certainly false
- I can't decide

Please, explain in only one sentence what made you decide (we need this to validate your HIT):

Almost certainly true

Likely to be true

Likely to be false

Almost certainly false

Examples credible/not-credible

- Users believed these

- US construction plunges 10%
- Tropical storm in the gulf
- Markets down on Portugal, Greece
- Yankees' Bob Sheppard dies

- Users did not believe these

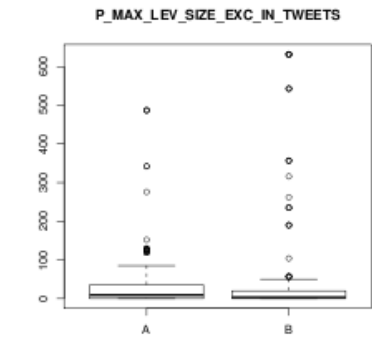
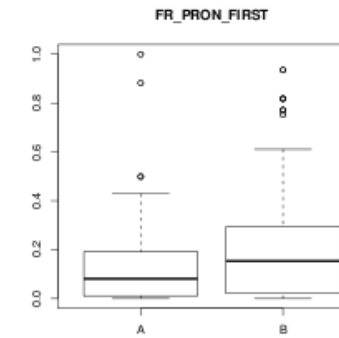
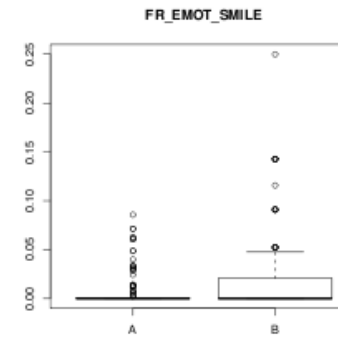
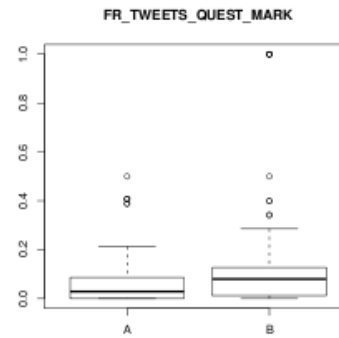
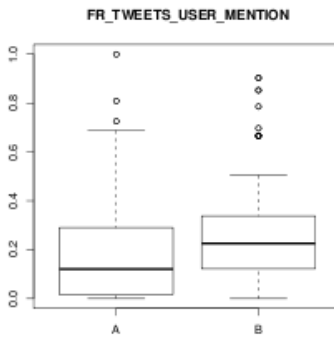
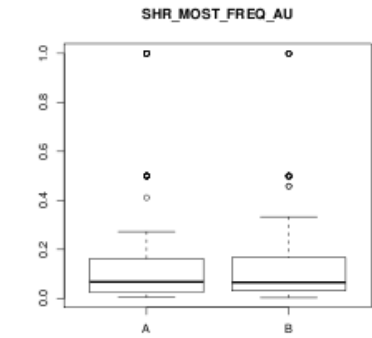
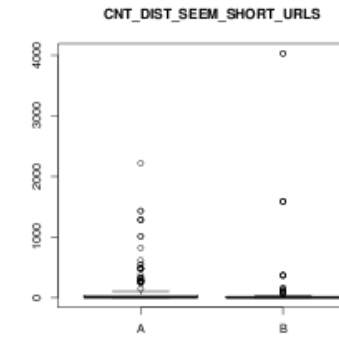
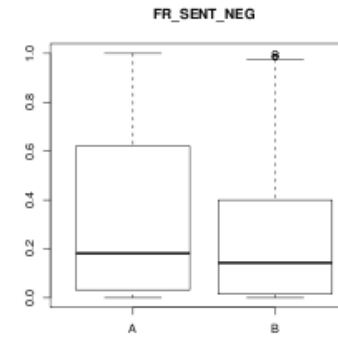
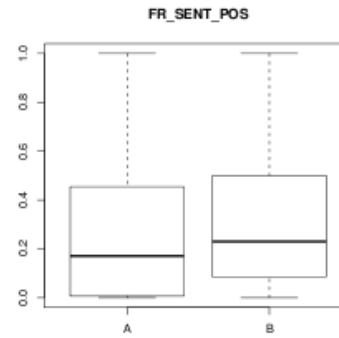
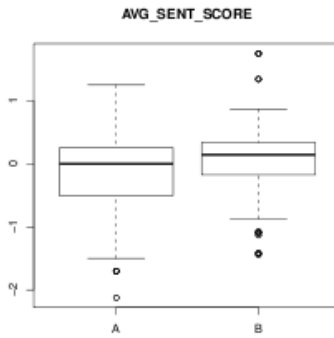
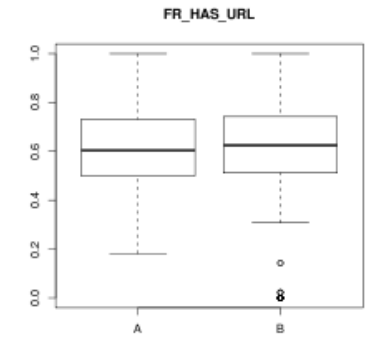
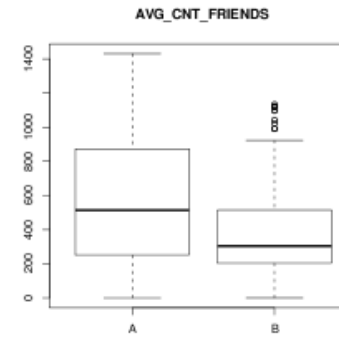
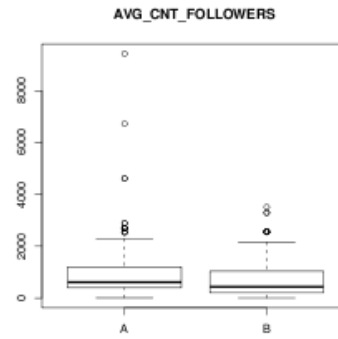
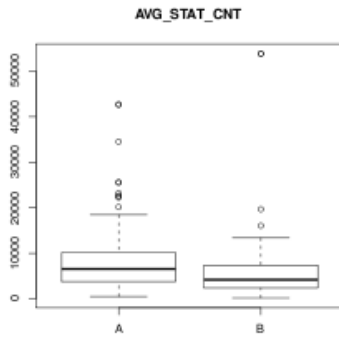
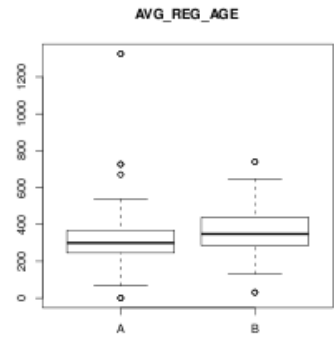
- Lennon lyrics auctioned for 1.2m (true)
- Suspicious vehicle causes NYC scare (false alarm: no bomb was found, they were gas canisters)
- Free video calling app for Android (spam)
- Warnings on BART murder trial verdict (partially true: no widespread riots)

Credible tweets for users tend to ...

- Have a URL
- Don't have exclamation marks
- Express a negative sentiment
- Are re-posted by prolific users
- Are re-posted by well-connected users

A = credible

B = non-credible



Text subset					
Class	TP Rate	FP Rate	Prec.	Recall	F ₁
A	0.636	0.152	0.808	0.636	0.712
B	0.848	0.364	0.700	0.848	0.767
W. Avg.	0.742	0.258	0.754	0.742	0.739
Network subset					
A	0.667	0.212	0.759	0.667	0.71
B	0.788	0.333	0.703	0.788	0.743
W. Avg.	0.727	0.273	0.731	0.727	0.726
Propagation subset					
A	0.606	0.091	0.870	0.606	0.714
B	0.909	0.394	0.698	0.909	0.789
W. Avg.	0.758	0.242	0.784	0.758	0.752
Top-element subset					
A	0.727	0.152	0.828	0.727	0.774
B	0.848	0.273	0.757	0.848	0.800
W. Avg.	0.788	0.212	0.792	0.788	0.787

4 features related to the share of the top-1:
 URL, user mention, author, hashtag

Summary

- Detected past events on twitter
- Took a sample of 10 tweets per event
- Asked users if they believe the tweets
- Built an automatic classifier
- Got 86% accuracy on balanced classes