Understanding Social Media

Ricardo Baeza-Yates

Alejandro Jaimes

Yahoo! Research, Barcelona

Social Media

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

Social Media

- Relationships are not
  - Discrete
  - Fixed
  - Explicit

Social Media

- 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

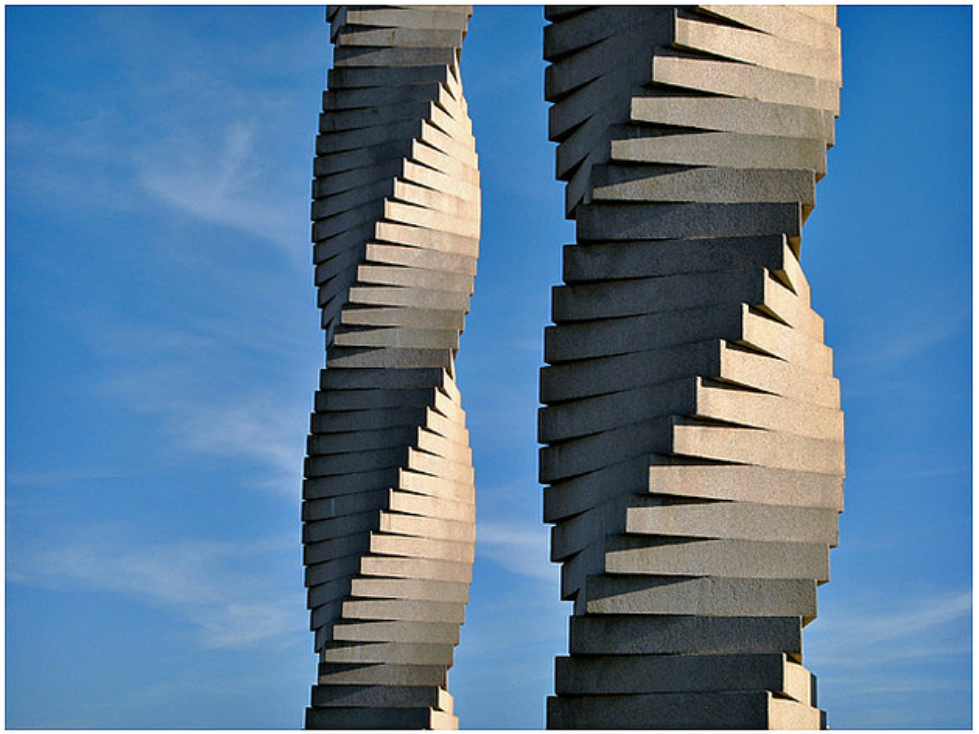
Placing Flickr Photos on a Map

flickr from YAHOO!

You aren't signed in Sign In Help

Home The Tour Sign Up Explore Upload Search

Favorite Actions Share this ← Newer Older →



By pasaro
Pablo Saludes Rodil

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

129 views 2 comments 1 favorite

Columnas UAB

Presentada a DPChallenge

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)

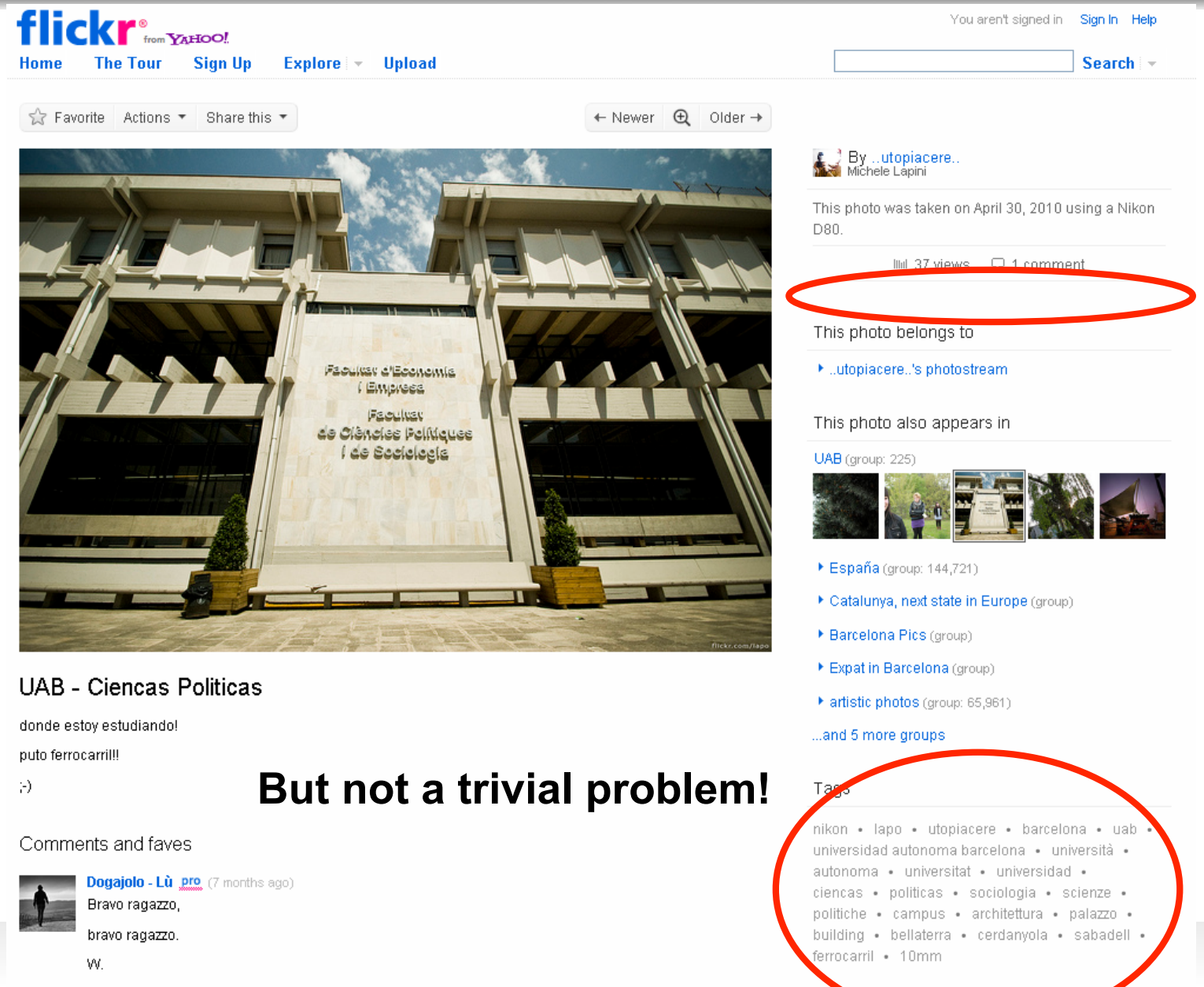
...and 5 more groups

Comments and faves

Bellprat (32 months ago)
Sembla que ballin

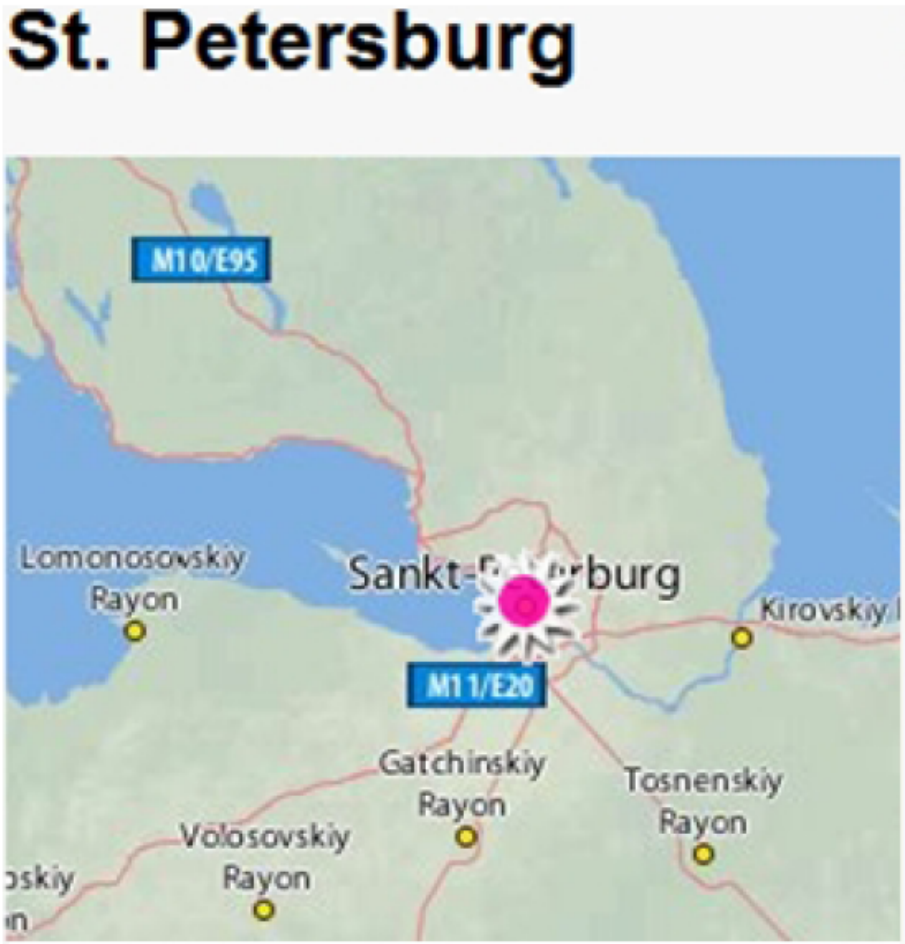
Tags

Placing Flickr Photos on a Map

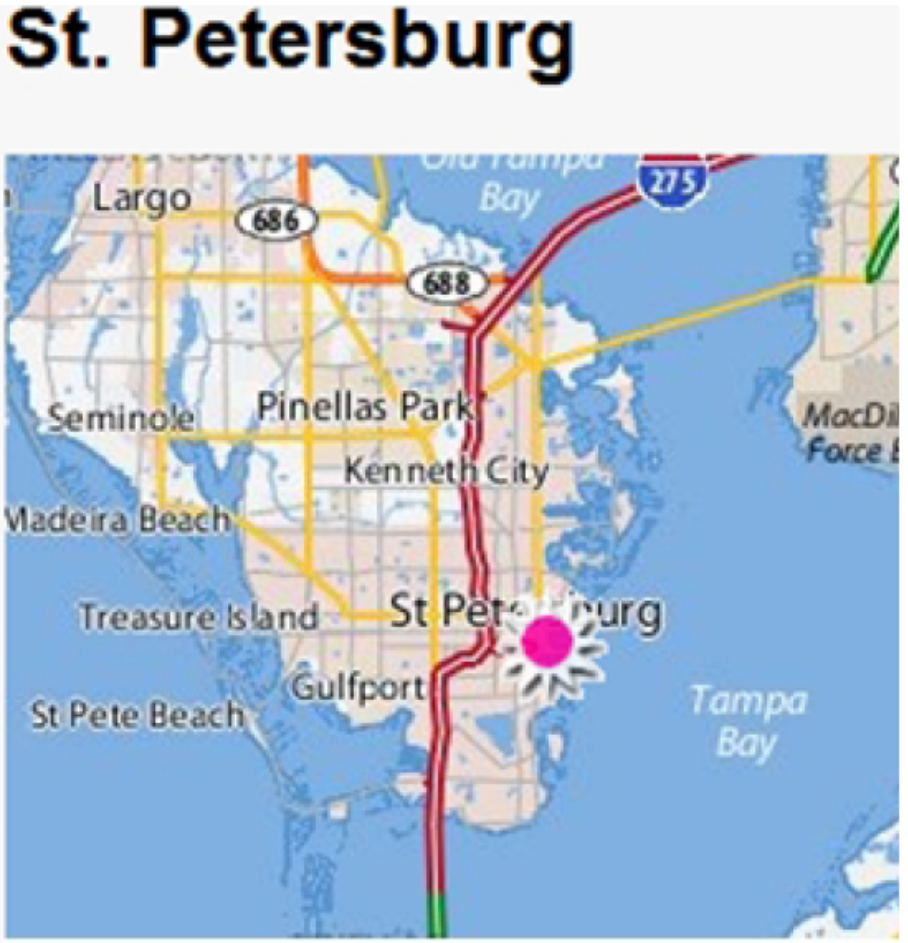


But not a trivial problem!

Placing Flickr Photos on a Map



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



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

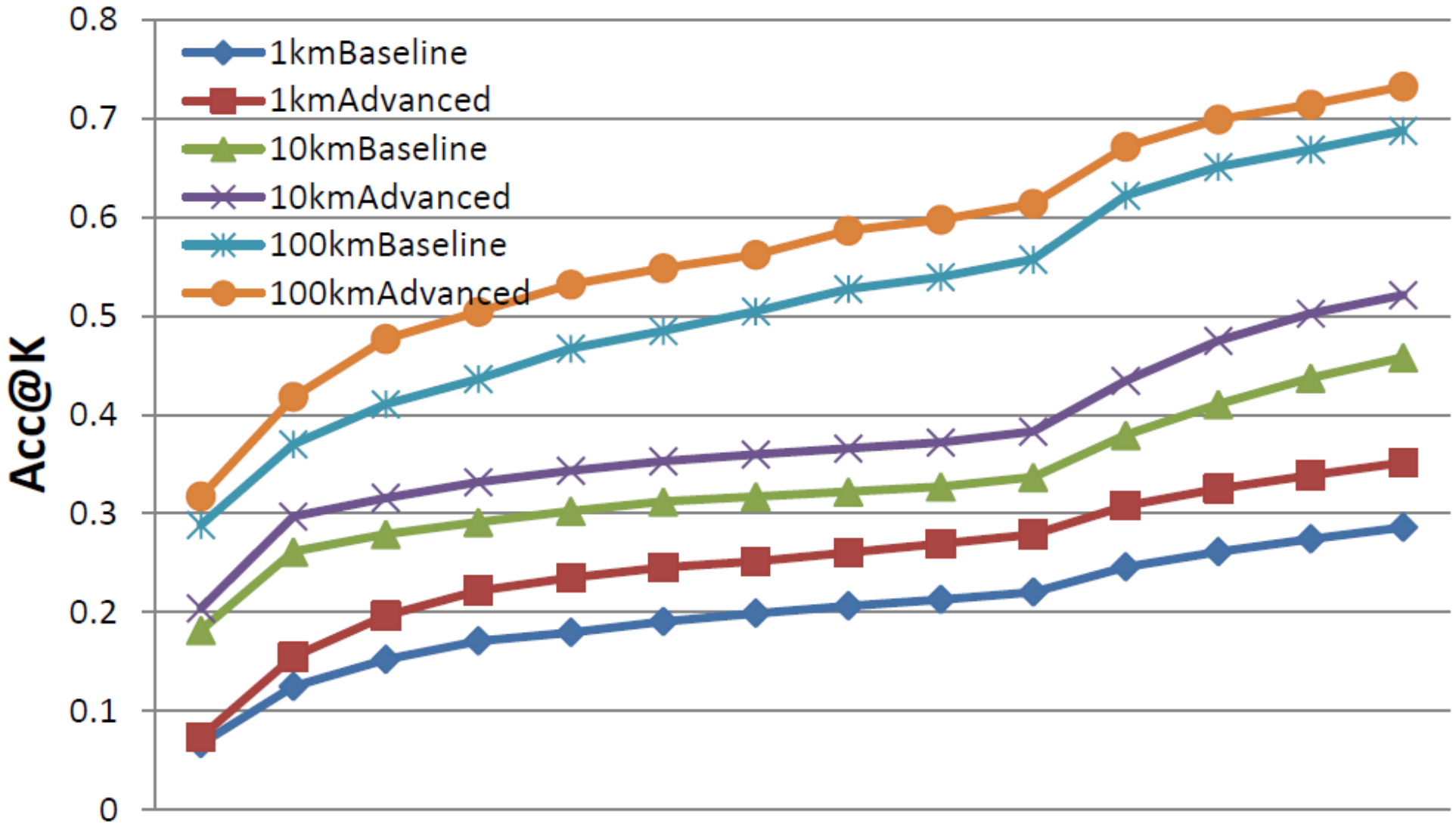
Sparsity!

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

“smoothing”

Work by Vanessa Murdock and his team

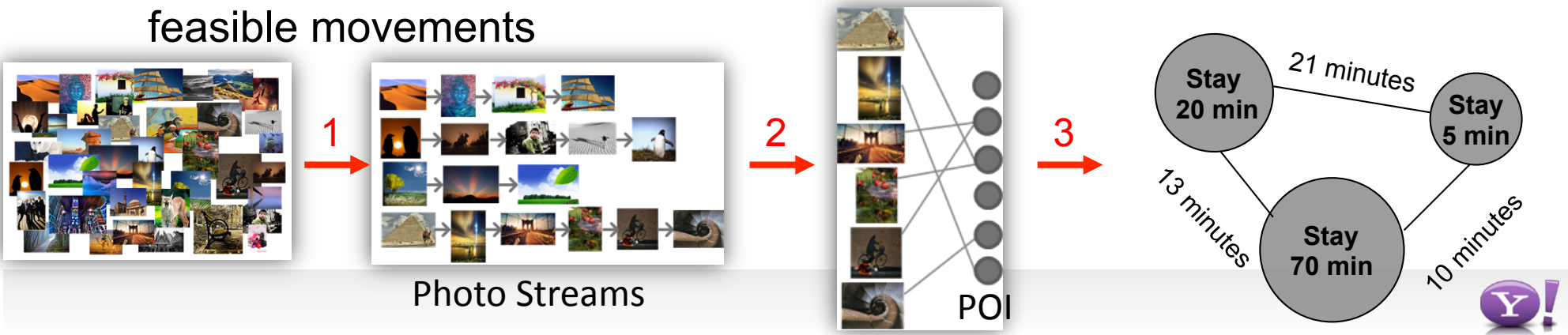
Placing Flickr Photos on a Map



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



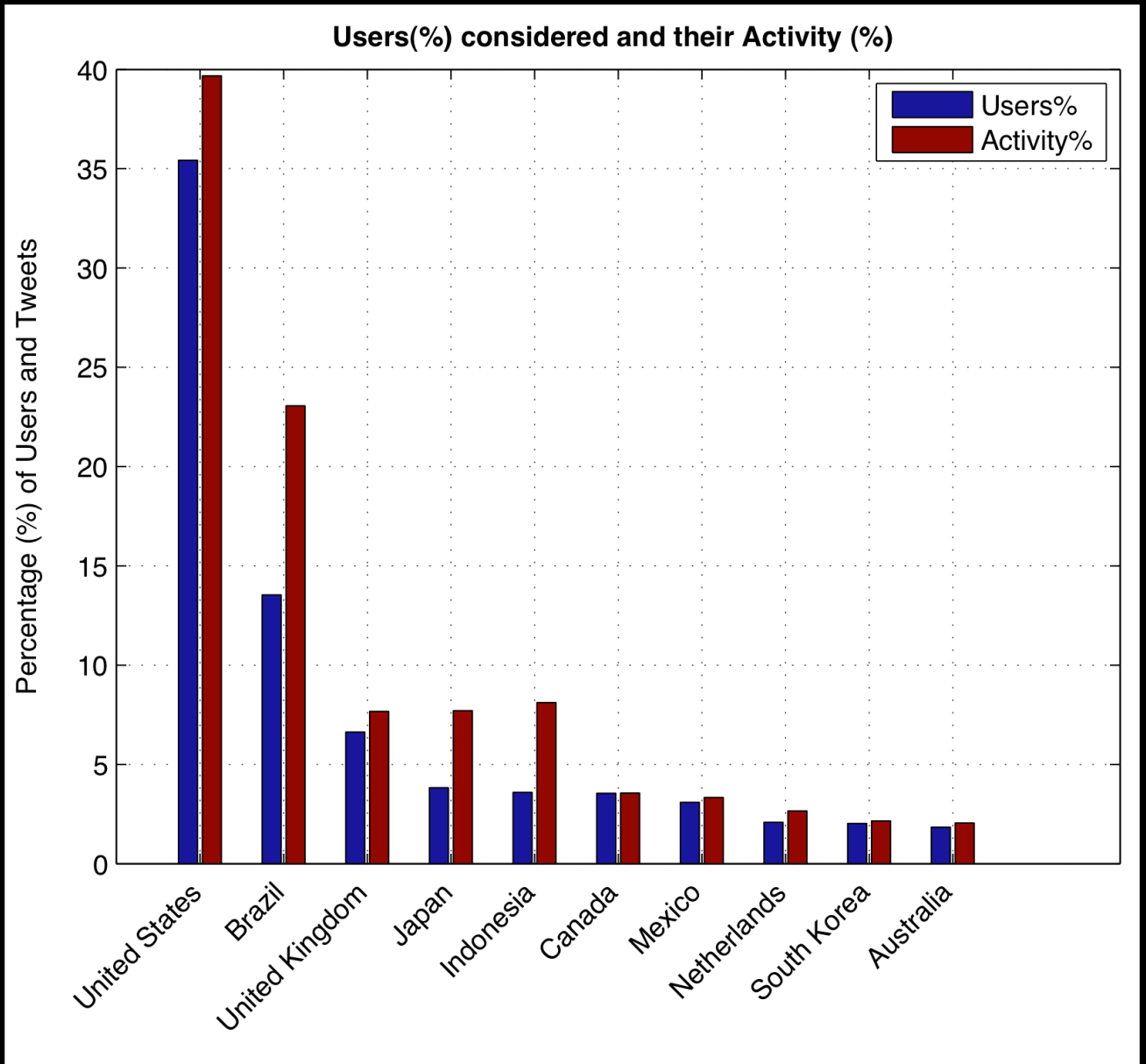
Microblogging Behavior

[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

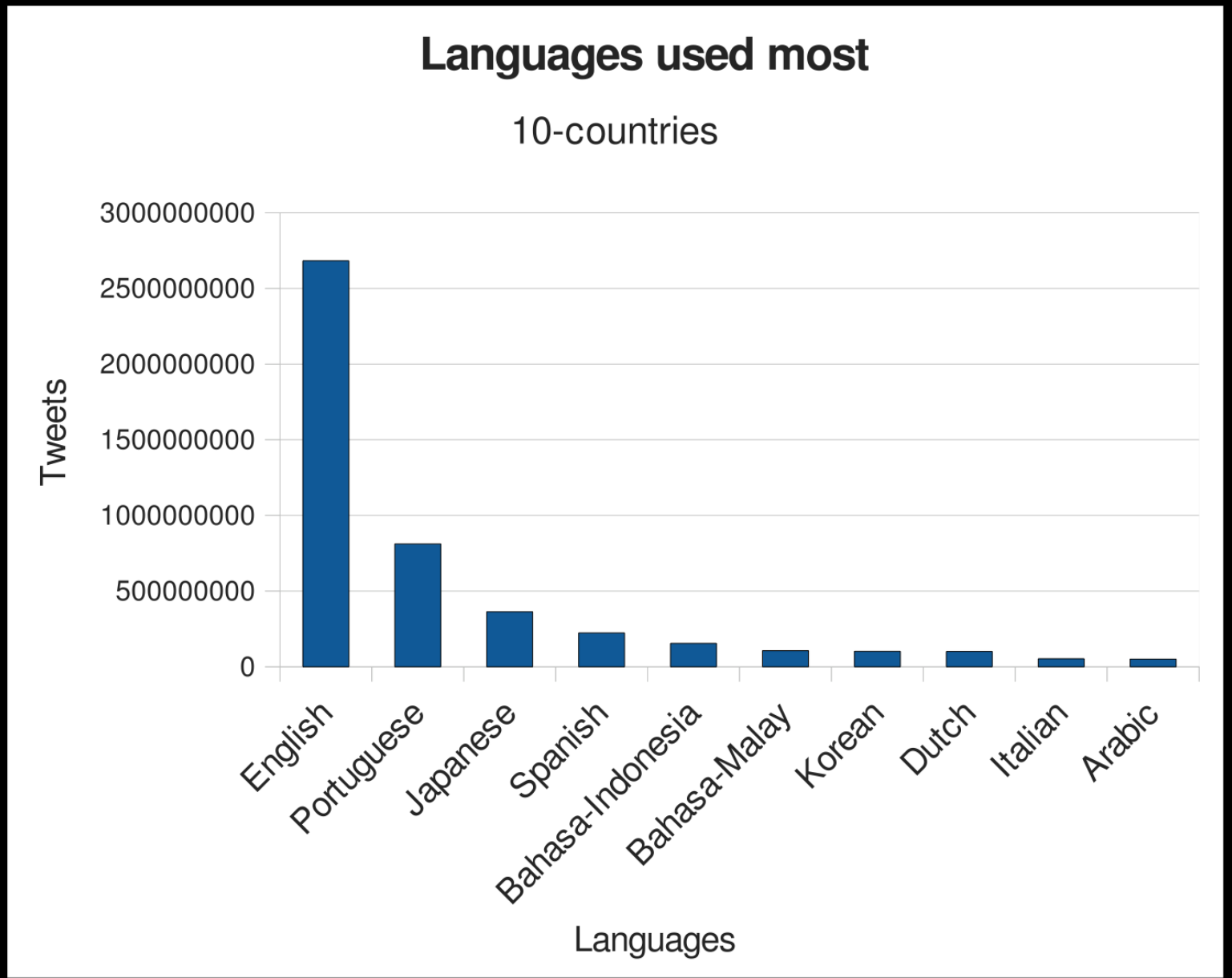
Microblogging Behavior

Results I



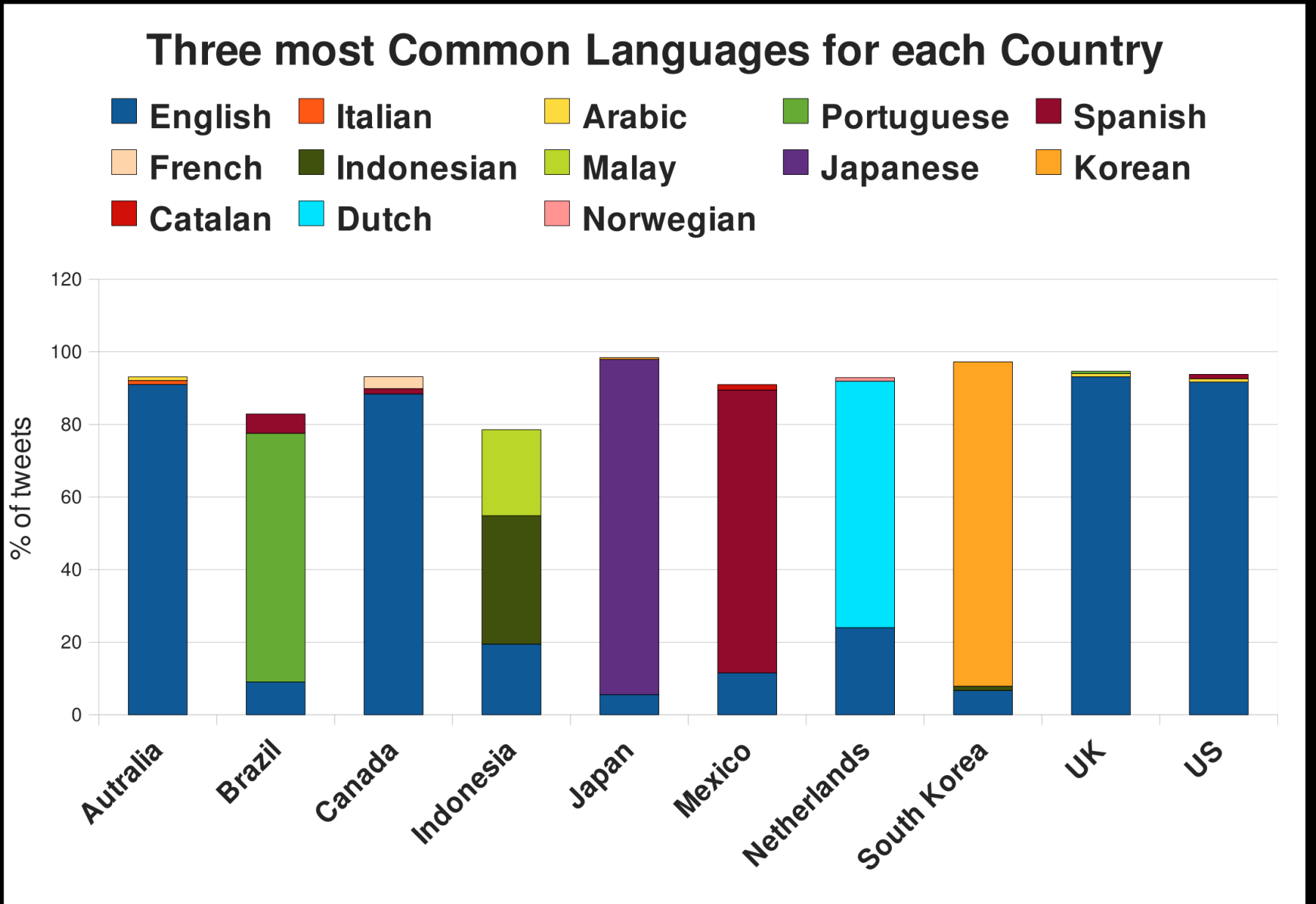
Microblogging Behavior

Results II



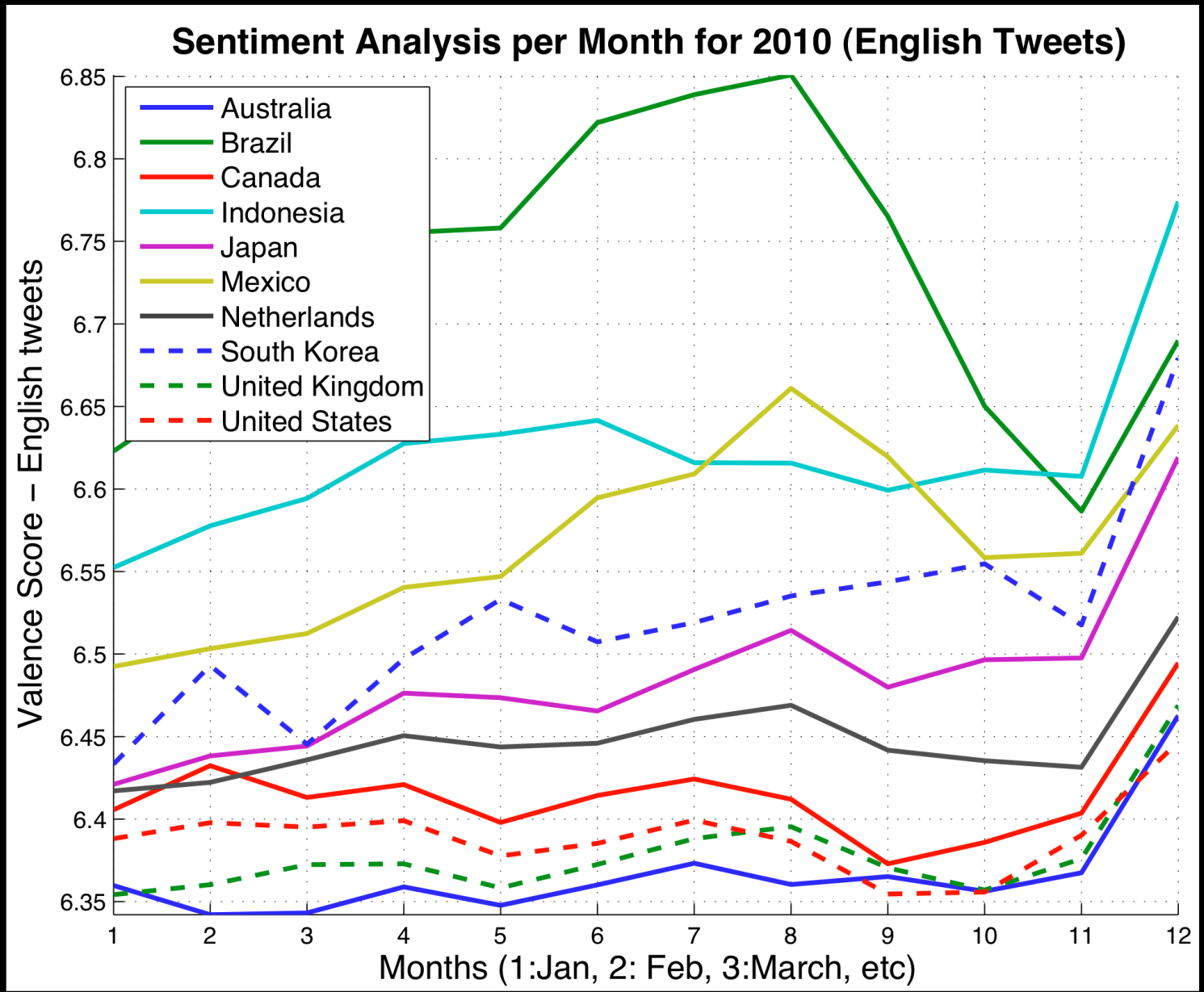
Microblogging Behavior

Results III



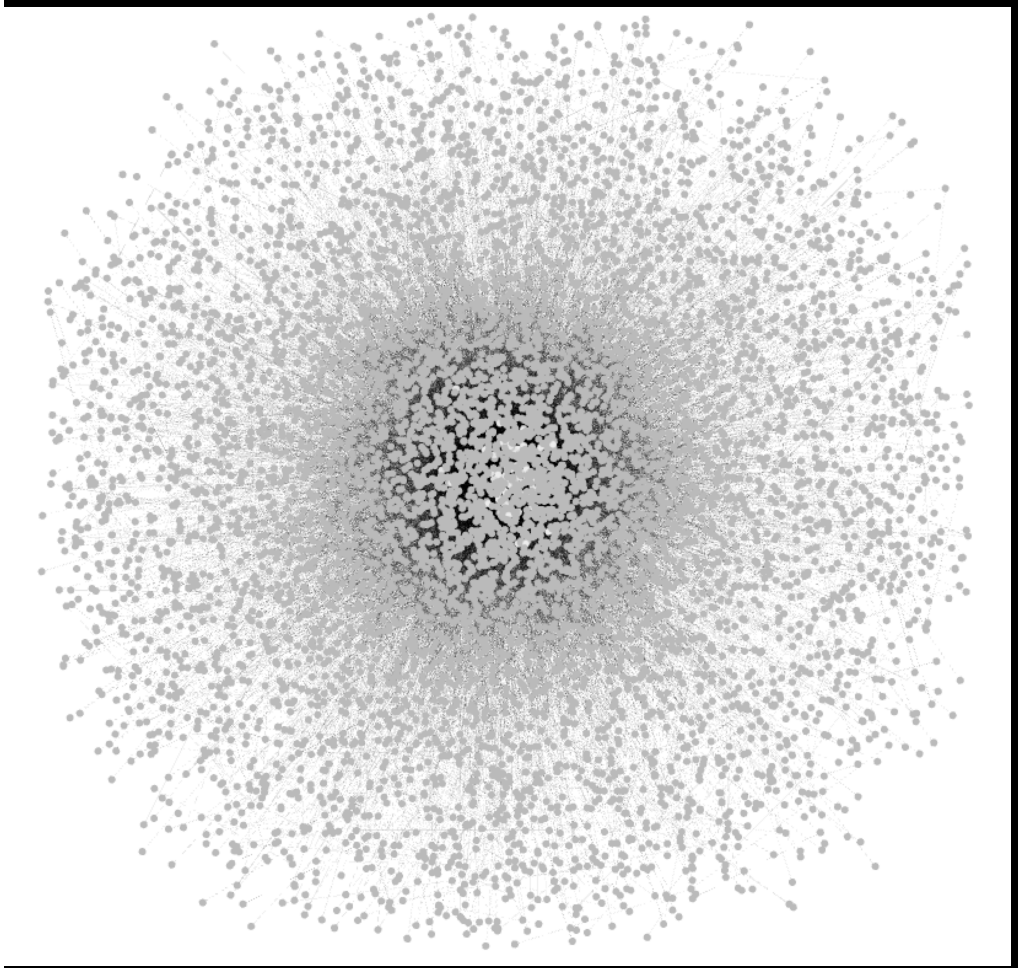
Microblogging Behavior

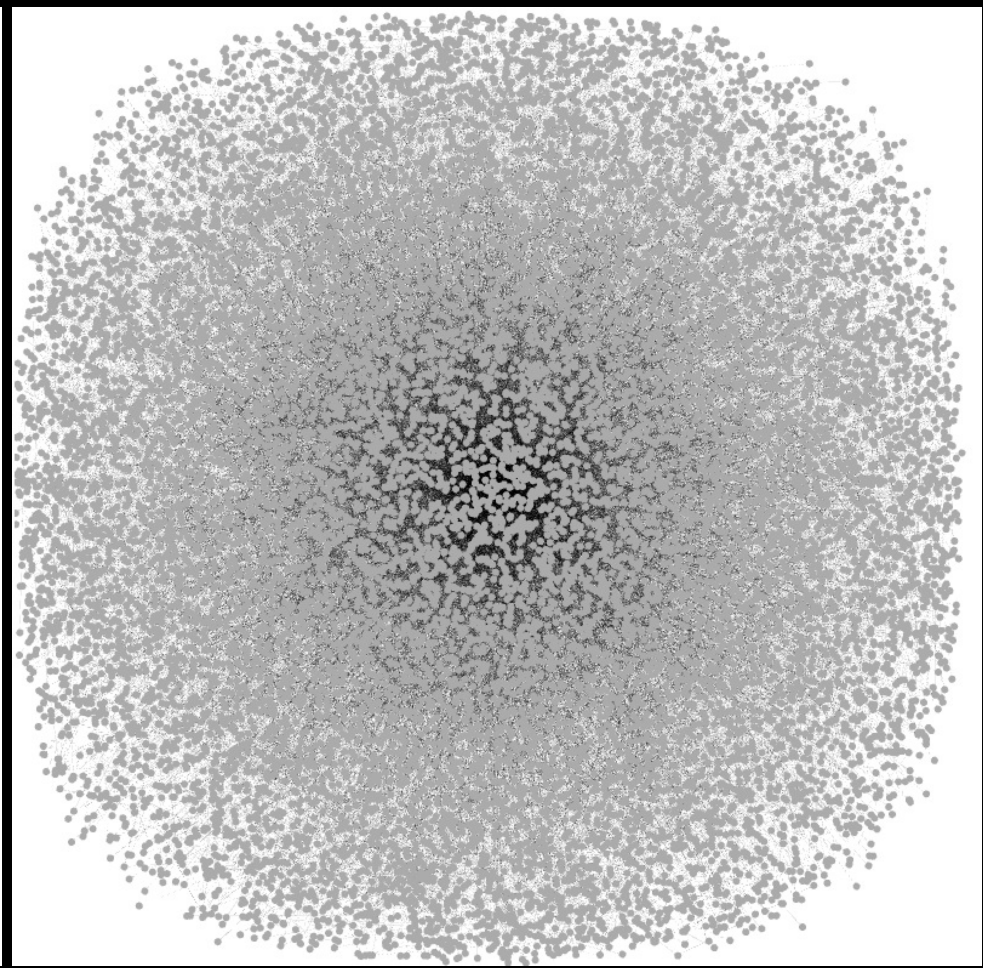
Results IV



Microblogging Behavior

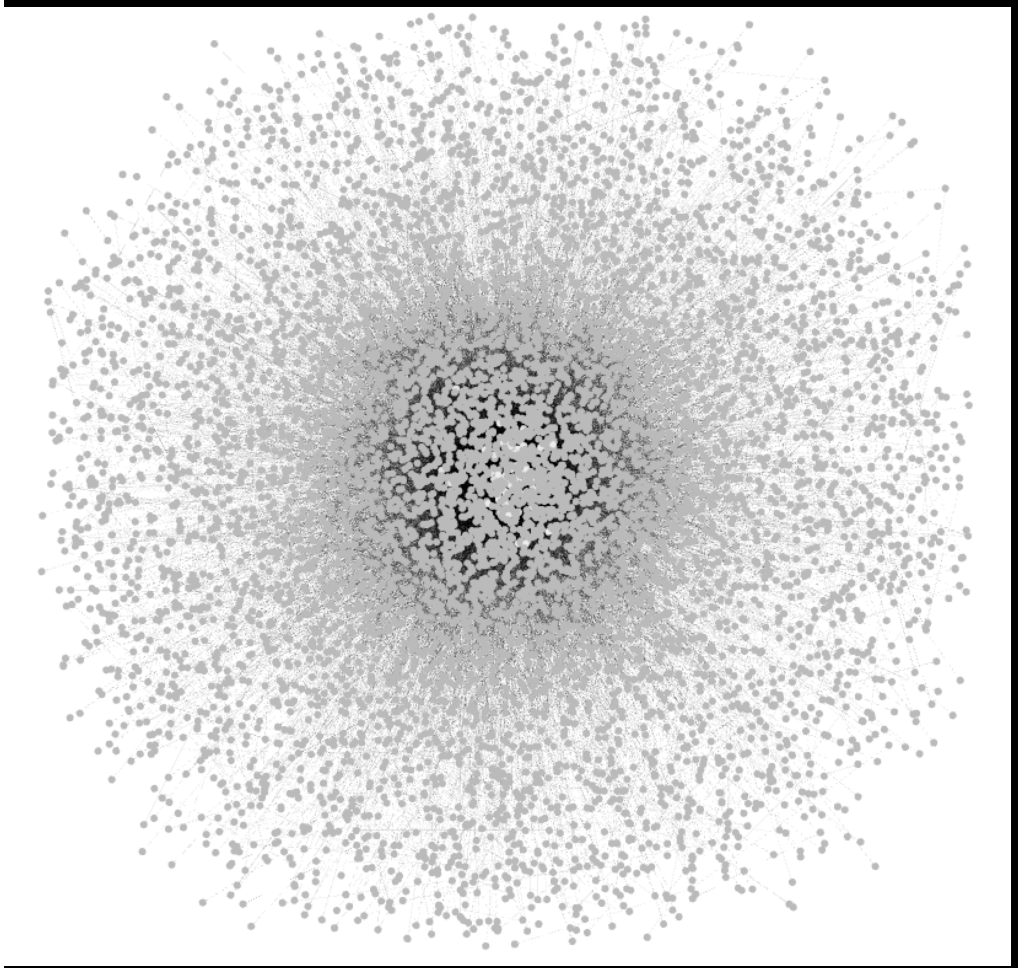
Can you guess which is Australia and which is Indonesia?



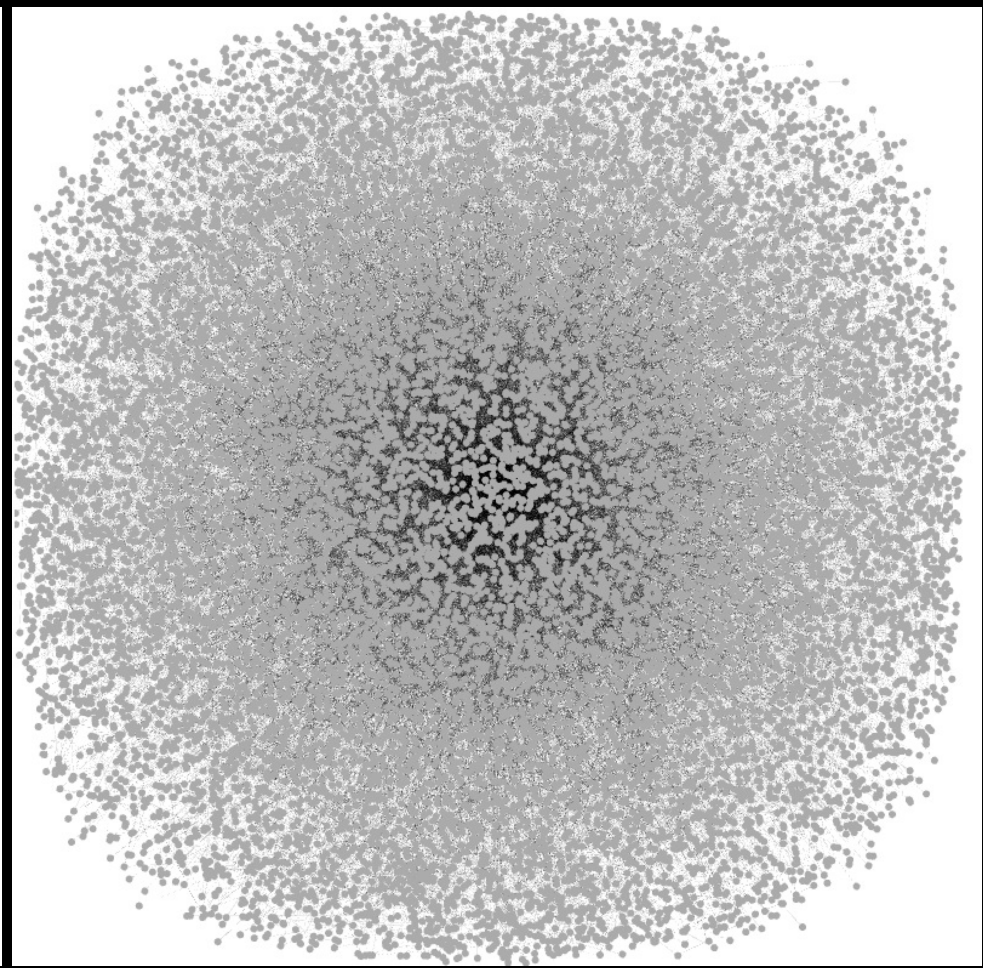


Microblogging Behavior

Australia

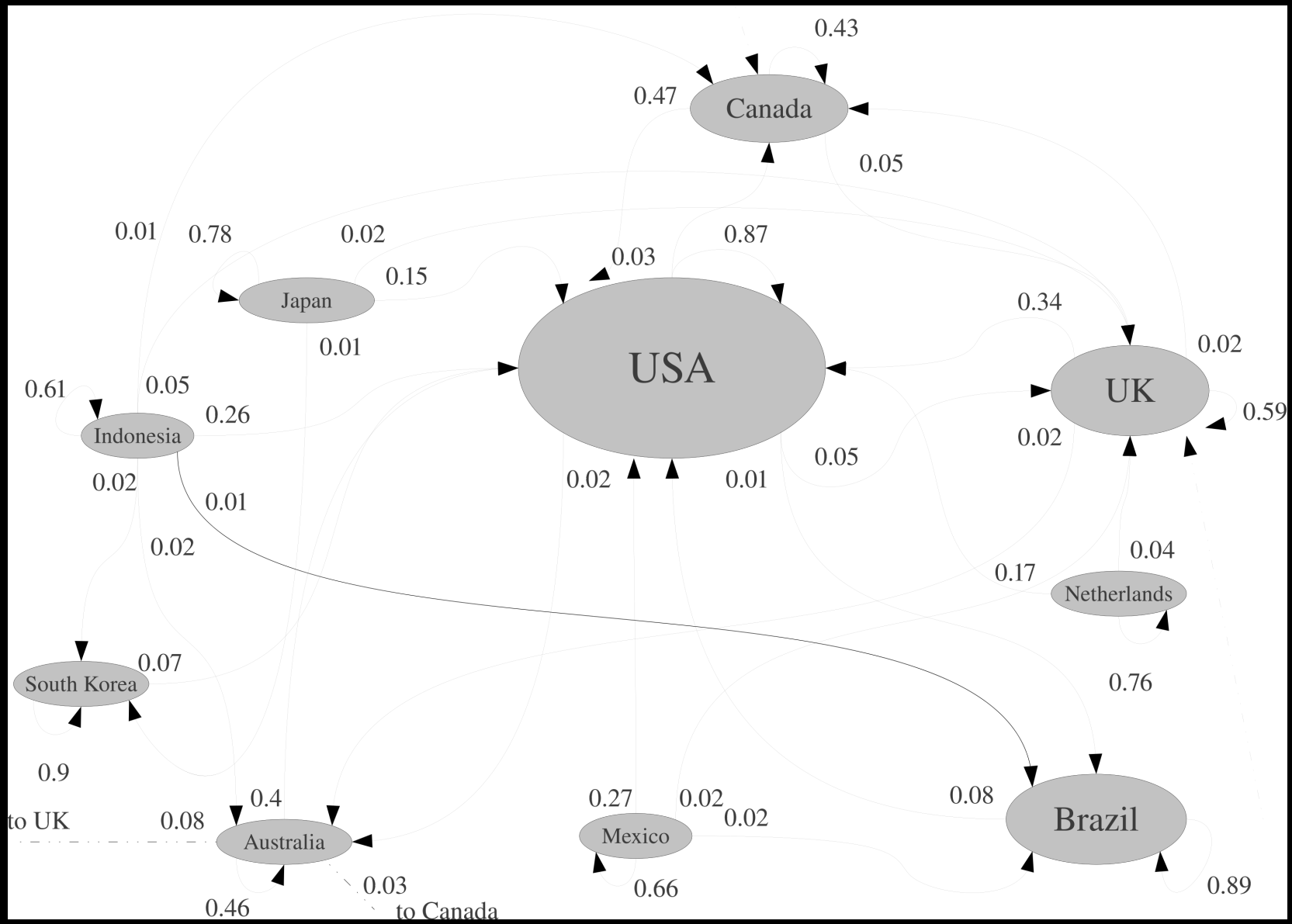


Indonesia



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

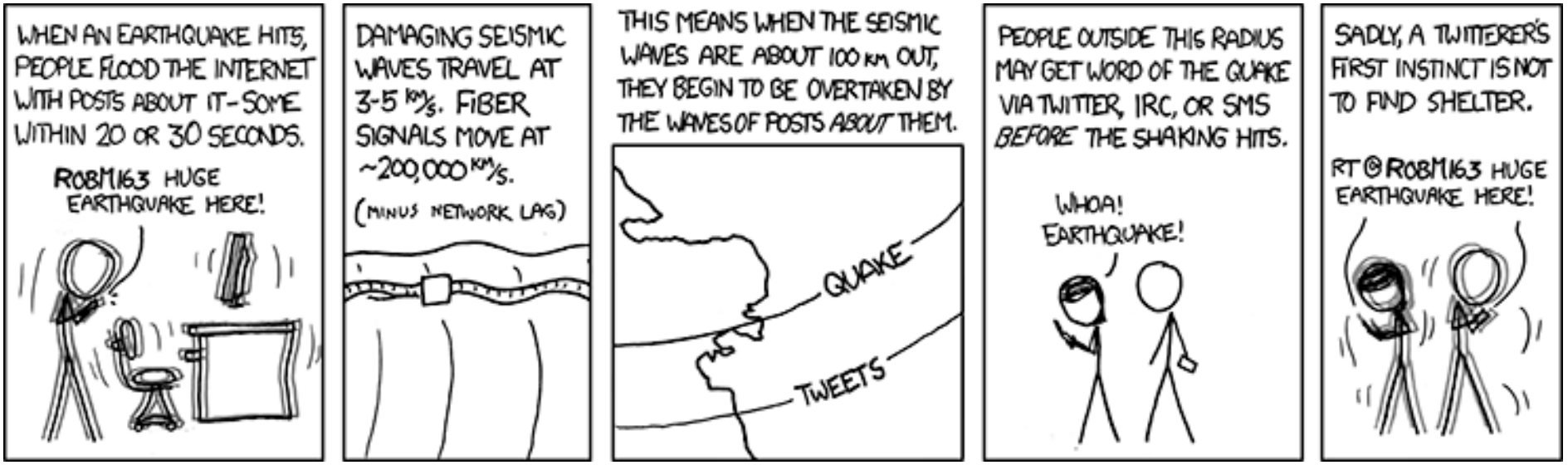
Twitter under Crisis

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

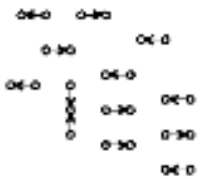
- Sat Feb 27, 2010. 03:34 local time
- 7th largest recorded in history
  - 8.8 Mw, 90 s, 500-600 casualties
  - Haiti: 7.7 Mw, 60 s, 90k-300k casualties
- Communications
  - Almost impossible for 2-3 hours
  - First video images 6-7 hours after quake

MASACRITICA.Blogspot.com
27febrero10

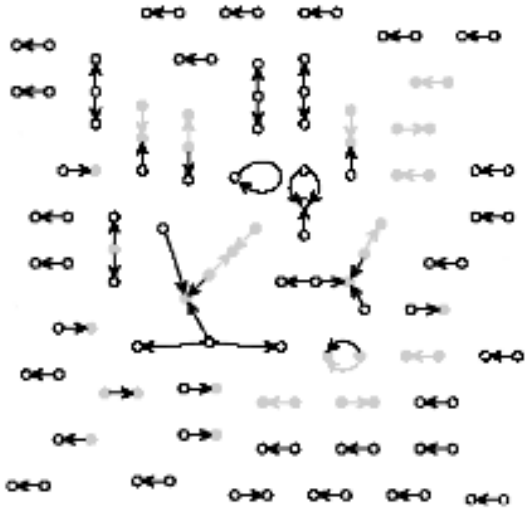
Chileans love Twitter

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

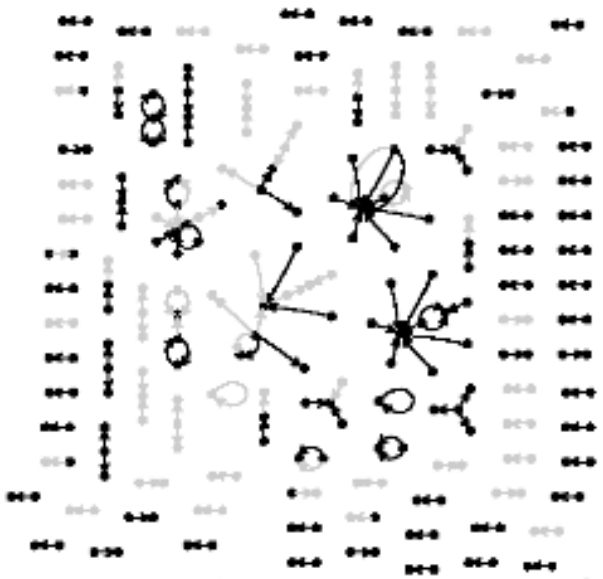
twitter



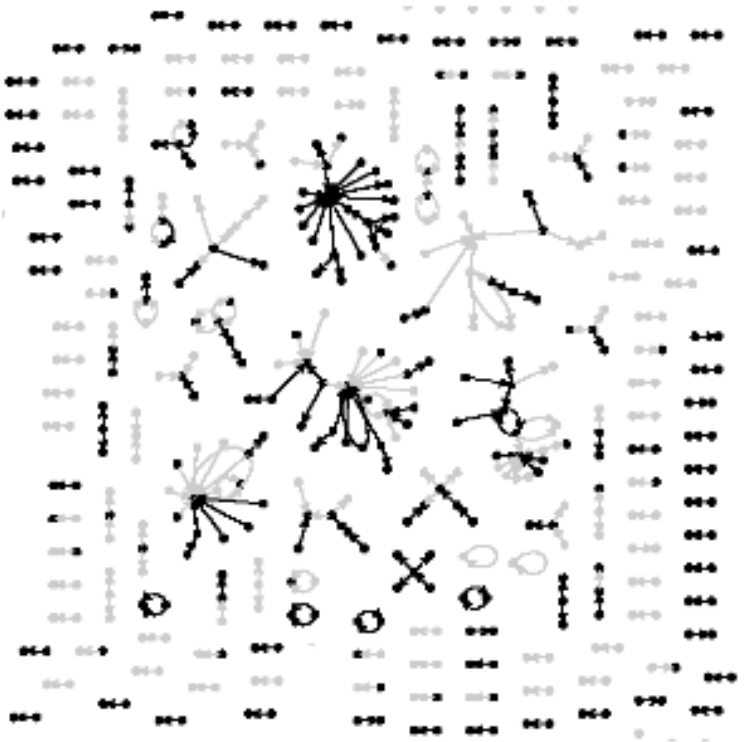
(a) 03:35:00 - 03:49:00



(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

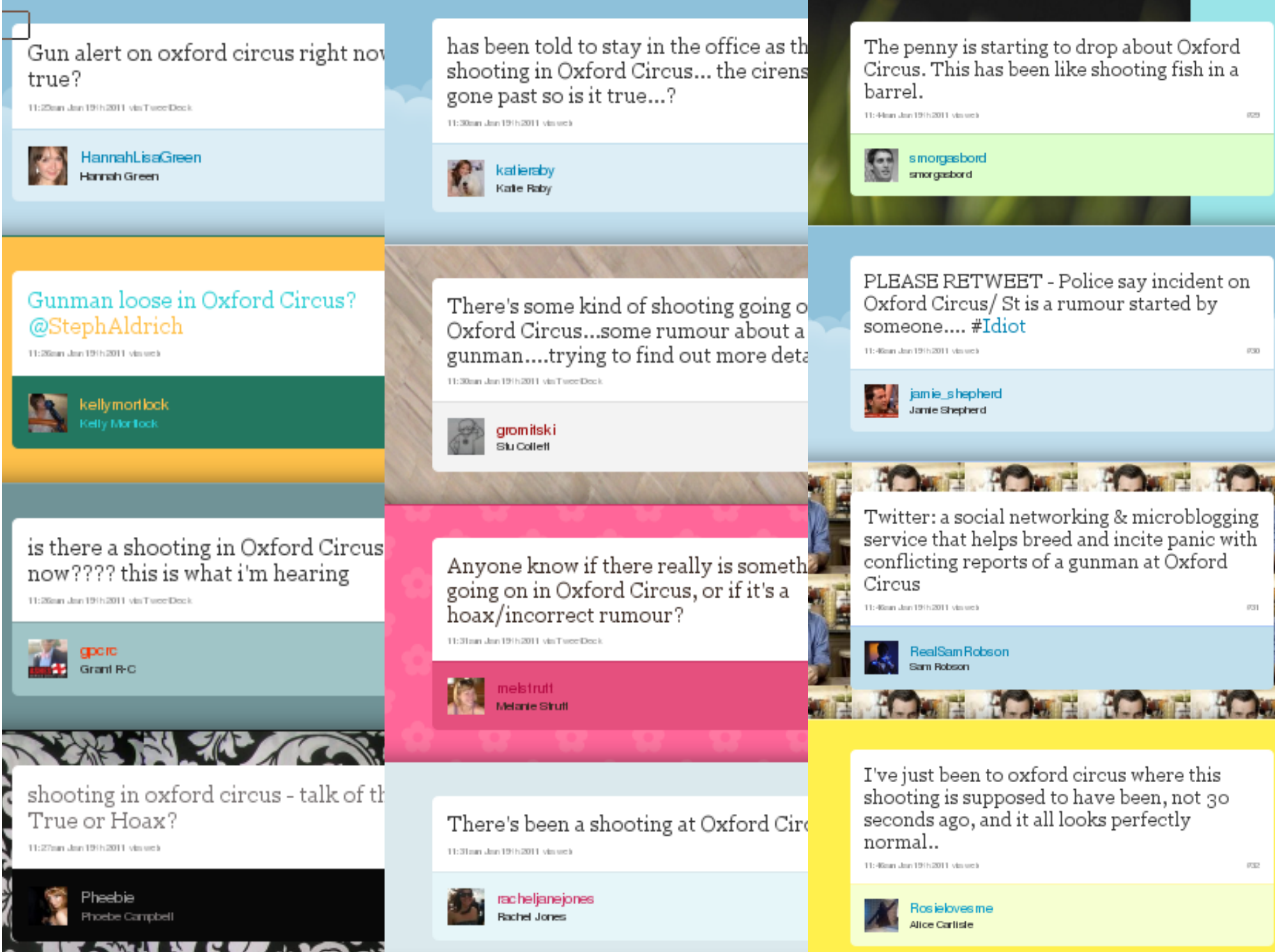
Case	# of unique tweets	% of re-tweets	# of unique "affirms"	# of unique "denies"	# of unique "questions"
Confirmed tru					
The internation ...tiago is closed	301	21	291	0	7
The *Viña del Mar International Song Festiva*			256	0	3
Fire in the Chemistry Faculty at the University of Concepcion	42	49	38	0	4
Navy acknowledges mistake informing about tsunami warning				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					
Death of artist Ricardo Arjona	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

TRUE

FALSE

How many people confirmed?

How many people denied?


Gun alert on oxford circus right nov
true?
HannahLisaGreen
Hannah Green
has been told to stay in the office as th
shooting in Oxford Circus... the cirens
gone past so is it true...?
katieraby
Katie Raby
The penny is starting to drop about Oxford Circus. This has been like shooting fish in a barrel.
smorgasbord
smorgasbord
Gunman loose in Oxford Circus? @StephAldrich
kellymortlock
Kelly Mortlock
There's some kind of shooting going o
Oxford Circus...some rumour about a
gunman....trying to find out more deta
gromitski
Stu Collett
PLEASE RETWEET - Police say incident on Oxford Circus/ St is a rumour started by someone.... #Idiot
jamie_shepherd
Jamie Shepherd
is there a shooting in Oxford Circus
now???? this is what i'm hearing
gpcrc
Grant R-C
Anyone know if there really is someth
going on in Oxford Circus, or if it's a
hoax/incorrect rumour?
melstrutt
Melanie Strutt
Twitter: a social networking & microblogging service that helps breed and incite panic with conflicting reports of a gunman at Oxford Circus
RealSamRobson
Sam Robson
shooting in oxford circus - talk of th
True or Hoax?
Pheebie
Phoebe Campbell
There's been a shooting at Oxford Circ
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..
Rosielovesme
Alice Carlisle

Can we automatically detect false tweets?

Can we automatically detect false tweets?

Of course not.

Information credibility

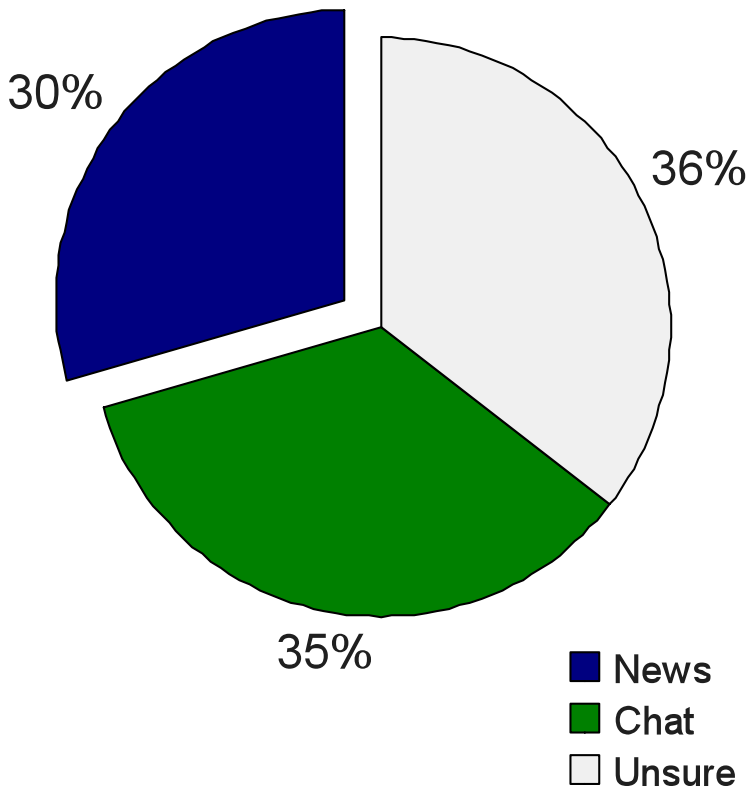
- Perceived quality
- Made of multiple dimensions
- Examples:
  - Same article written by male/female author
  - Same headline in news/twitter
  - ...

Supervised classification

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

Labeling: News or Chat

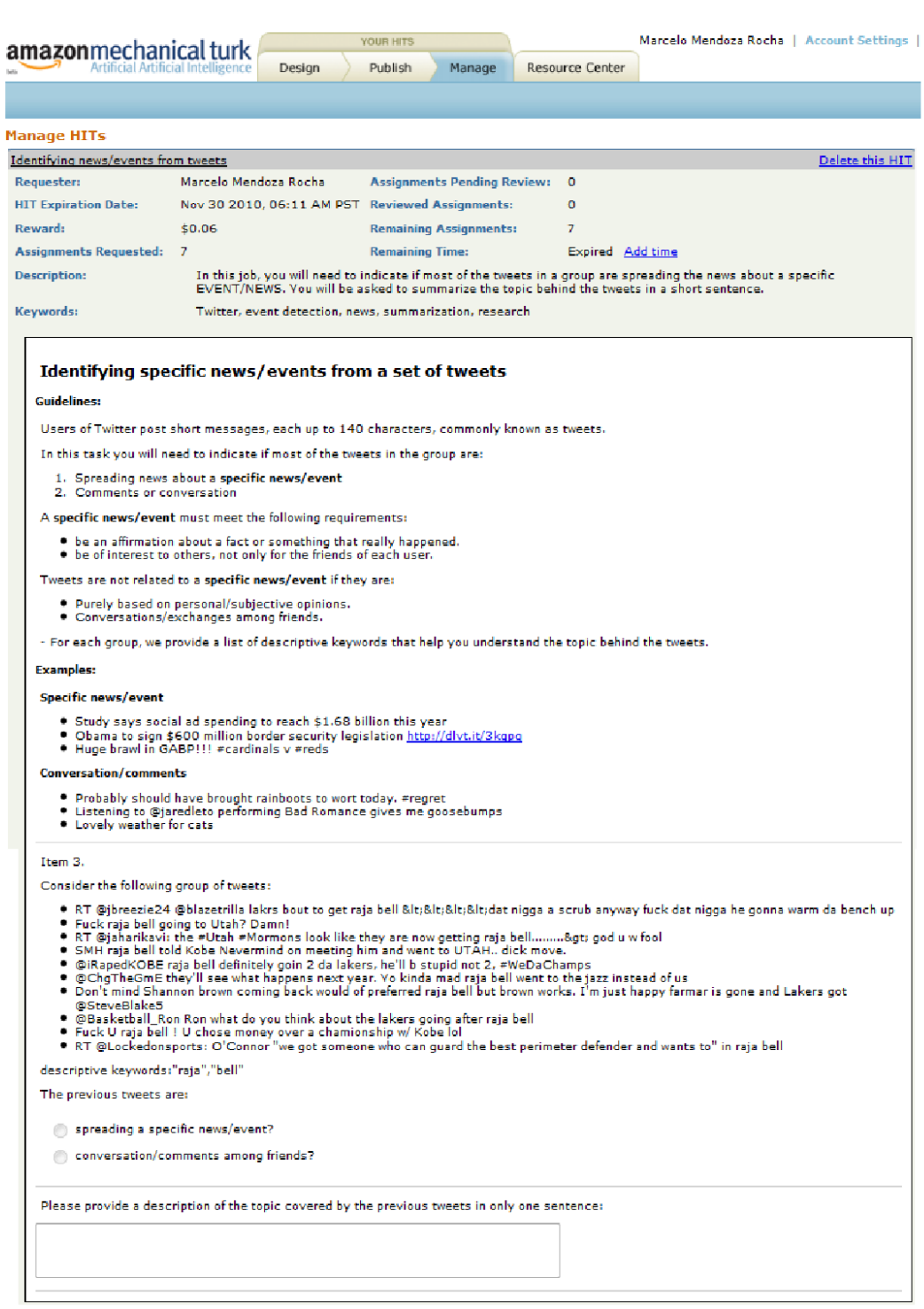
- 383 events from TwitterMonitor.net [Mathioudakis & Koudas]
- 7 evaluators per event
- >=5 agreement



Spreading a specific news/event

-OR-

Conversation or comments among friends.



amazonmechanical turk Artificial Artificial Intelligence

YOUR HITS: Design | Publish | Manage | Resource Center

Marcelo Mendoza Rocha | Account Settings |

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
Reward:	$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 sentance.

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://dlvr.it/3kqpq
- 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 @jbreezie24 @blazetrilla lakrs bout to get raja bell <<<<dat nigga a scrub anyway fuck dat nigga he gonna warm da bench up
- Fuck raja bell going to Utah? Damn!
- RT @jaharikavi: the #Utah #Mormons look like they are now getting raja bell.....> god u w fool
- SMH raja bell told Kobe Nevermind on meeting him and went to UTAH.. dick move.
- @iRapedKOBE 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 farmar is gone and Lakers got @SteveBlake5
- @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"

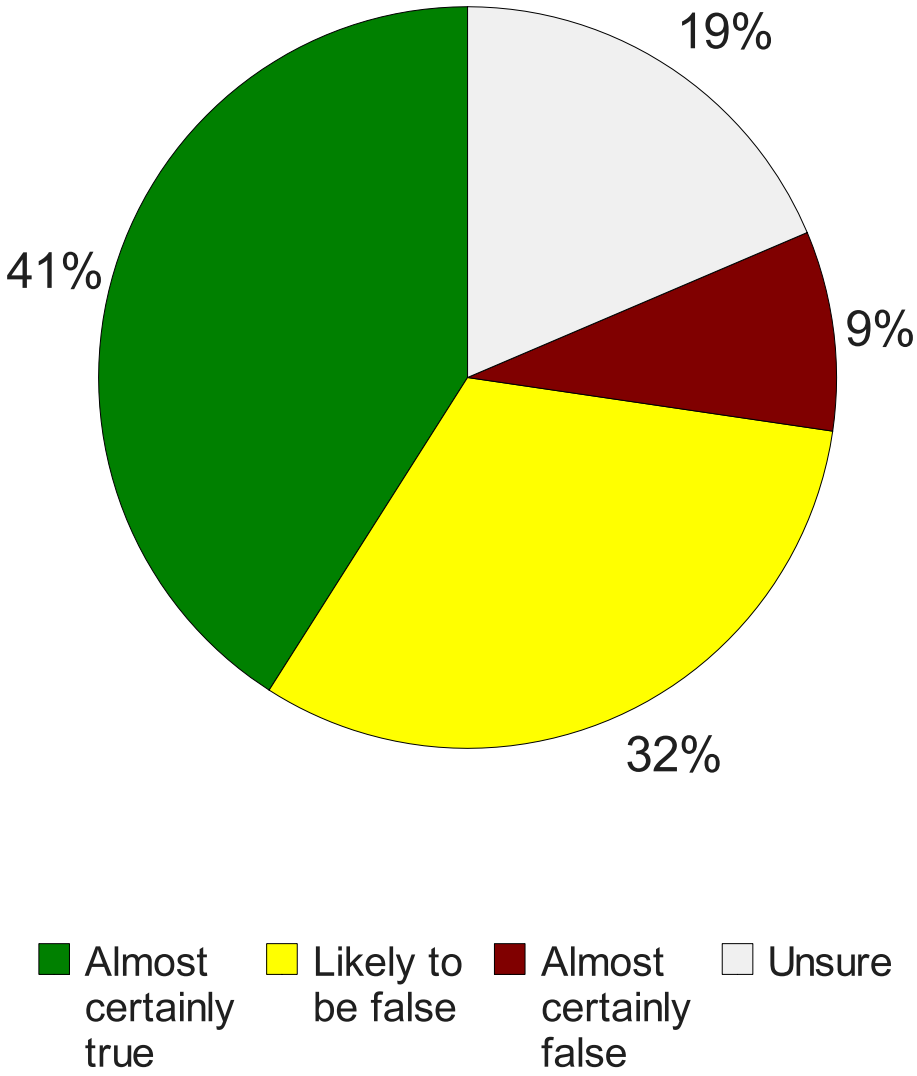
The previous tweets are:

- ( ) 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:

Labeling: Credible or Not

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

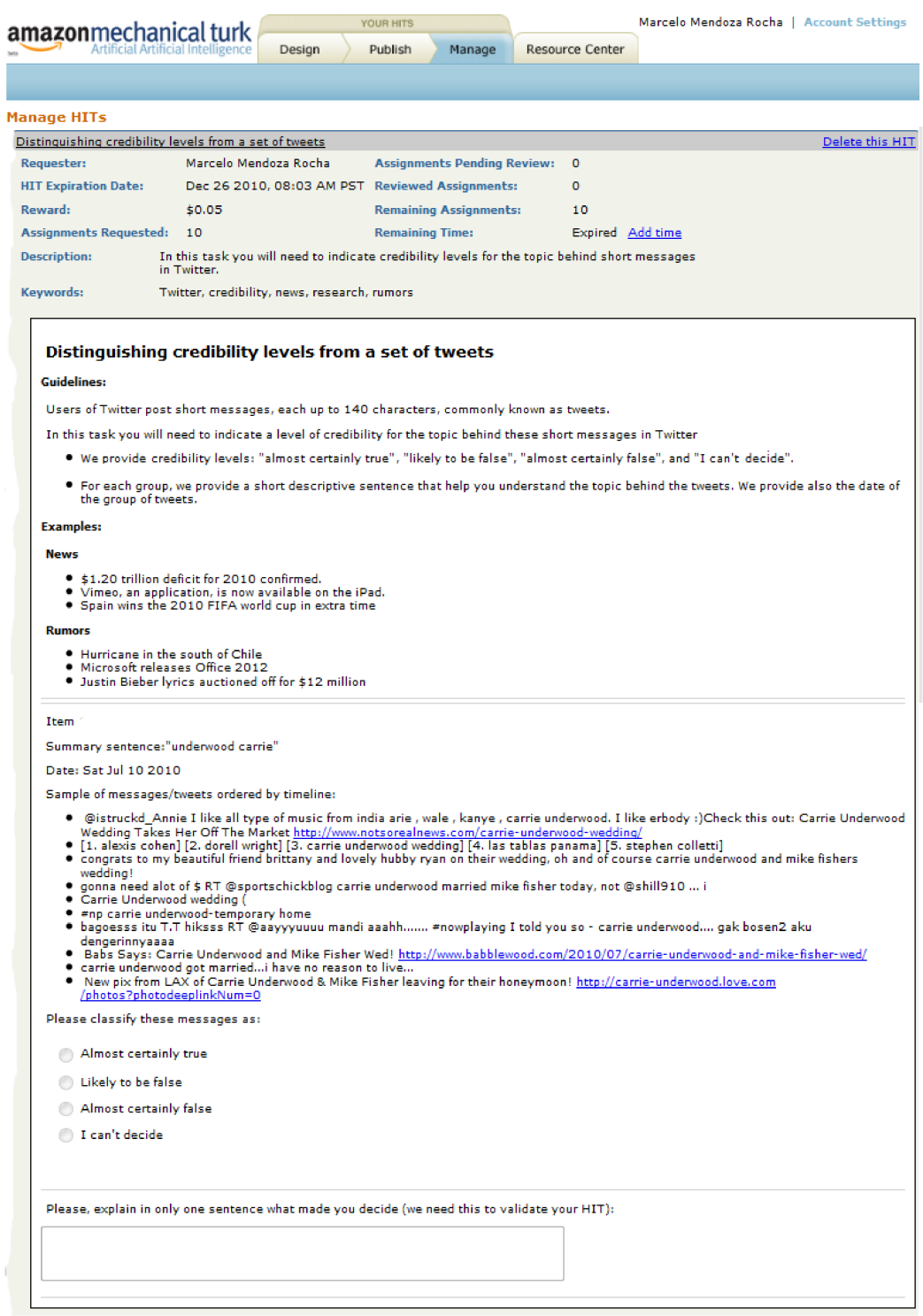


Almost certainly true

Likely to be true

Likely to be false

Almost certainly false



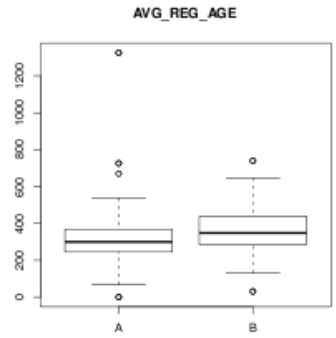
Examples credible/not-credible

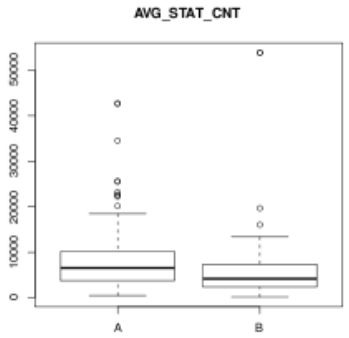
- Users believed these
  - US construction plunges 10%
  - Tropical storm in the gulf
  - Markets down on Portugal, Greece
  - Yankees' Bob Sheppard dies
- Users did not believe these
  - Lennon lyrics auctioned for 1.2m (true)
  - Suspicious vehicle causes NYC scare (false alarm: no bomb was found, they were gas canisters)
  - Free video calling app for Android (spam)
  - Warnings on BART murder trial veredict (partially true: no widespread riots)

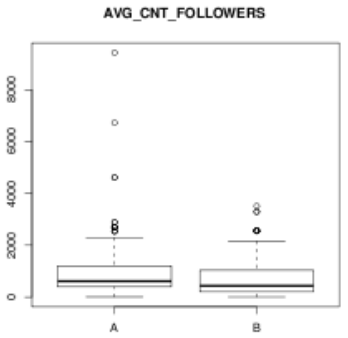
Credible tweets for users tend to ...

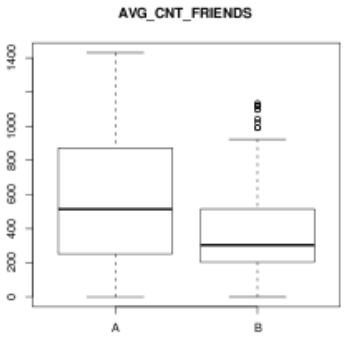
- Have a URL
- Don't have exclamation marks
- Express a negative sentiment
- Are re-posted by prolific users
- Are re-posted by well-connected users

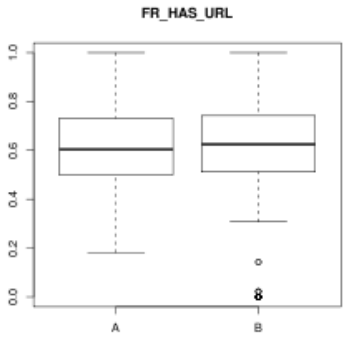
A = credible B = non-credible

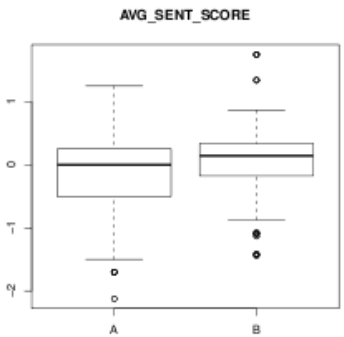


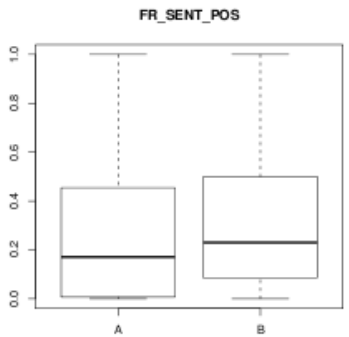


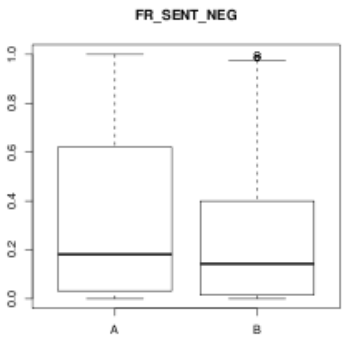


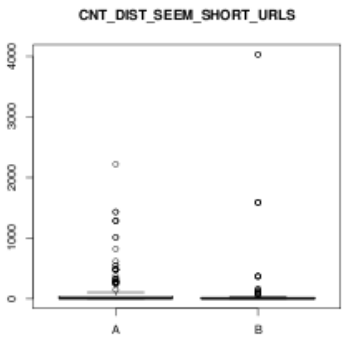


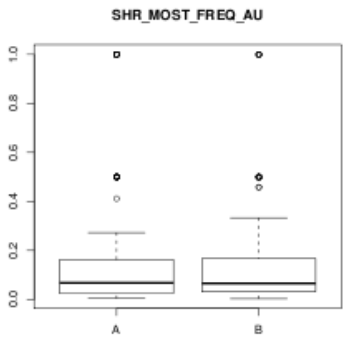


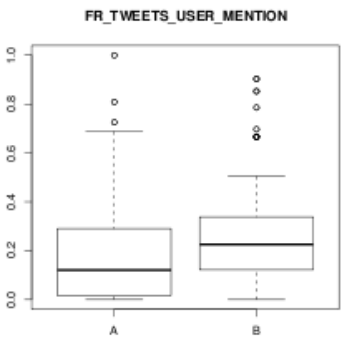


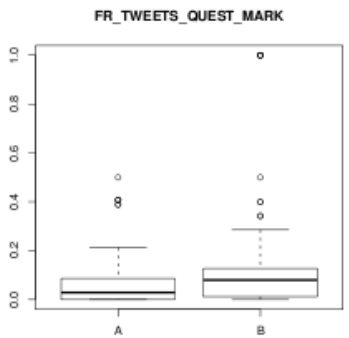


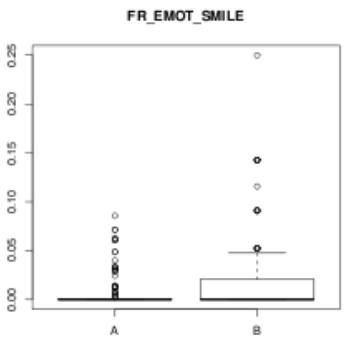


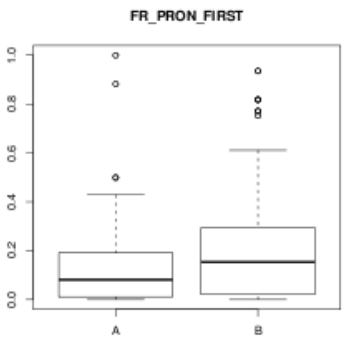


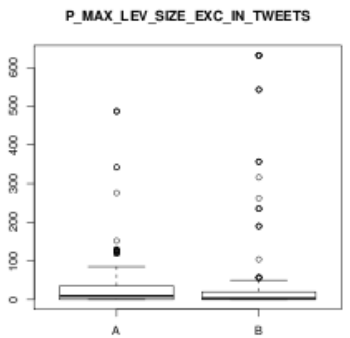












Class	TP Rate	FP Rate	Prec.	Recall	F_1
Text subset					
A	0.636	0.152	0.808	0.636	0.712
B	0.848	0.364	0.700	0.848	0.767
W. Avg.	0.742	0.258	0.754	0.742	0.739
Network subset					
A	0.667	0.212	0.759	0.667	0.71
B	0.788	0.333	0.703	0.788	0.743
W. Avg.	0.727	0.273	0.731	0.727	0.726
Propagation subset					
A	0.606	**0.091**	**0.870**	0.606	0.714
B	**0.909**	0.394	0.698	**0.909**	0.789
W	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