Spatial Datasets And Urban Applications

AAAI ICWSM TUTORIAL, MAY 15, 2017 MONTREAL, CANADA

https://github.com/bmtgoncalves/ICWSM17

Theory Session:

- **8:30 - 9:30** - *Bruno Gonçalves* - We will start with an overview of geolocated APIs such as those of Twitter and Foursquare, using Python focusing on their similarities and differences. Applications covered will range from visualizing Twitter usage and language geography , Inter-City Mobility through airline connections and the Global Language Network.

- **9:30 - 10:30** - *Anastasios Noulas* - In this part of the tutorial attendees will be provided with an overview perspective on location-based technologies with a focus on urban applications. From human mobility models and theory, to neighborhood detection and finding an appropriate location for placing a shop or amenity attendees will be introduced to the mechanics and building blocks of a number of research works with a strong application orientation. Next, the spotlight will be put on newer advancements in the area of the so-call gig economy, introducing services like Uber and the mechanics and building blocks of a number of research works with a strong application orientation. Next, the spotlight will be put on newer advancements in the area of the so-call gig economy, introducing services like Uber and Airbnb, highlighting opportunities for research using corresponding transport and hospitality datasets.

- **10:30 - 11:00** *Coffee break*

- **11:00 - 12:00** - *Desislava Hristova* - Overview of the theory of social capital in cities and the social role of places. This will include an introduction to structural holes and open network structures in light of social capital and how this can be extended to places in cities. The role of social media will be discussed as a proxy for understanding urban development and gentrification in cities like London and New York. Concrete examples of measuring the digital footprints of wellbeing in cities and beyond will be provided alongside the theory of cultural capital and urban development.

- **12:00 - 13:30** - *Lunch*

Practical Session:

- **13:30 - 14:30** - *Bruno Gonçalves* - After an introduction on how to collect data using the Twitter and Foursquare APIs we will proceed to perform several simple analyses to illustrate different important techniques and libraries. Using the airline transportation network (from the Bureau of Transportation Statistics) as a starting point, we will build a simple model of large scale mobility within the United States.

- **14:30 - 15:30** - *Anastasios Noulas* - In this part of the tutorial Anastasios will first focus on foundational methods for spatial retrieval using Python. Code snippets that allow for the organisation of spatial data, the calculation of geographic distances as well as the retrieval of points within a radius, given a pair of coordinates, will be examined in detail. Next, the Networkx library will be employed for the analysis of networks of places, whereas the session will close with a demo showcase of Uber's deck.gl framework for visualisation of spatial data.

- **15:30 - 16:00** *Coffee break*

- **16:00 - 17:00** - *Desislava Hristova* - Understanding Urban Development and Gentrification from Social Media. Social media is notoriously biased towards tech-savvy younger populations but can we exploit this bias to better understand processes of gentrification and predict urban development in neighbourhoods? We will look at geo-social data from Twitter and Foursquare alongside public data about deprivation in neighbourhoods to understand the contradictions which arise when complex urban processes such as gentrification take place. The role of symbolic capital such as social and cultural capital and the different ways of quantifying it will be explored as a follow-up to their theoretical definition in the first part of the tutorial. This will include hands-on exploration of predicting housing prices in developing neighbourhoods using standard statistical methods in Python.

Online Social Networks and Mobility - Theory

Bruno Gonçalves

www.bgoncalves.com

2016

every MINUTE of the DAY

PRESENTED BY DOMO

Americans USE **18,264,840** MEGABYTES OF WIRELESS DATA

BUZZFEED USERS VIEW **159,380** PIECES OF CONTENT

SNAPCHAT USERS WATCH **6,944,444** VIDEOS

Netflix SUBSCRIBERS STREAM **86,805** HOURS OF VIDEO

GOOGLE RECEIVES **2,276,867** SEARCH QUERIES

YOUTUBE USERS SHARE **400** HOURS OF NEW VIDEO

Instagram USERS LIKE **2,430,555** POSTS

FACEBOOK MESSENGER USERS SHARE **216,302** PHOTOS

SIRI ANSWERS **99,206** REQUESTS

Amazon MAKES **$222,283** IN SALES

Tinder USERS SWIPE **972,222** TIMES

3,567,850 TEXT MESSAGES ARE SENT IN THE **U.S.**

Giphy SERVES **569,217** GIFS

TWITTER USERS SEND **9,678** EMOJI-FILLED TWEETS

Dropbox USERS UPLOAD **833,333** NEW FILES

THE WEATHER CHANNEL RECEIVES **13,888,889** FORECAST REQUESTS

All this technology is making us antisocial

Social Media

Mobile devices

Geolocated Activity

@bgoncalves

www.bgoncalves.com

JAN 2016

GLOBAL DIGITAL SNAPSHOT

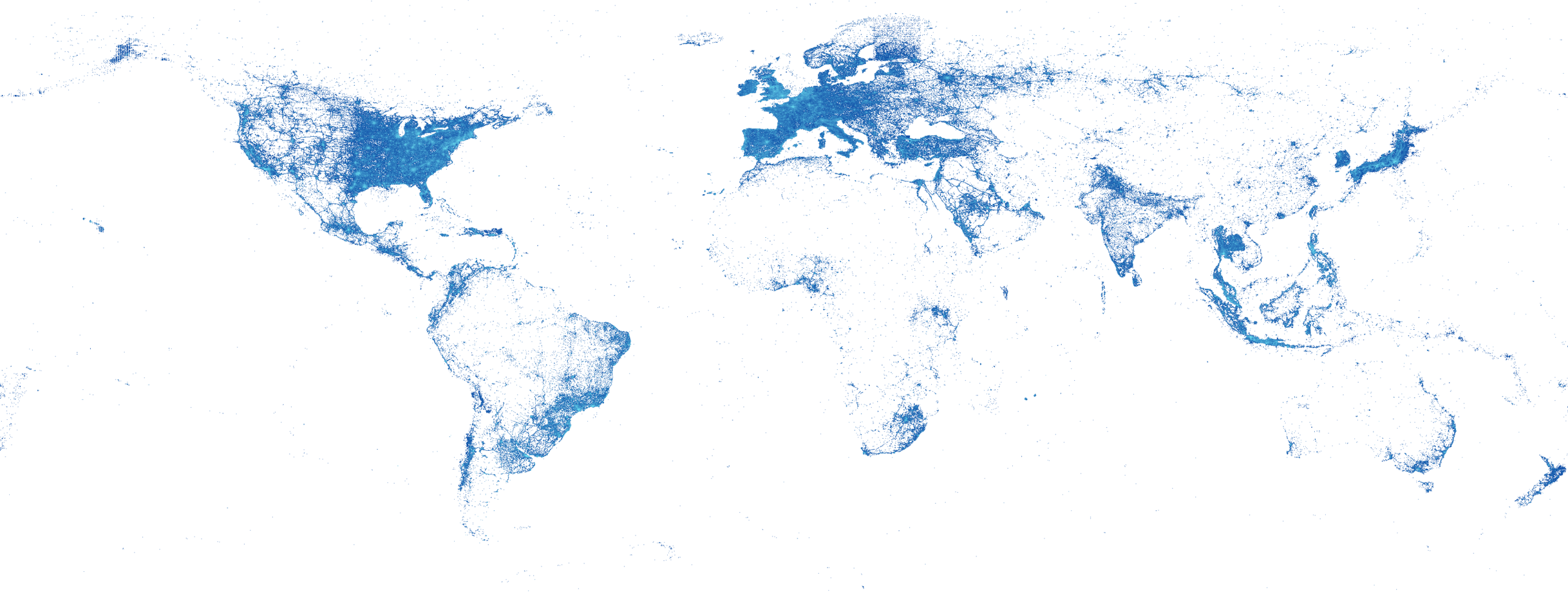
A SNAPSHOT OF THE WORLD'S KEY DIGITAL STATISTICAL INDICATORS

TOTAL POPULATION	INTERNET USERS	ACTIVE SOCIAL MEDIA USERS	UNIQUE MOBILE USERS	ACTIVE MOBILE SOCIAL USERS
7.395 BILLION	**3.419 BILLION**	**2.307 BILLION**	**3.790 BILLION**	**1.968 BILLION**
URBANISATION: 54%	PENETRATION: 46%	PENETRATION: 31%	PENETRATION: 51%	PENETRATION: 27%
FIGURE REPRESENTS TOTAL GLOBAL POPULATION, INCLUDING CHILDREN	*FIGURE INCLUDES ACCESS VIA FIXED AND MOBILE CONNECTIONS*	*FIGURE BASED ON ACTIVE USER ACCOUNTS, NOT UNIQUE INDIVIDUALS*	*FIGURE REPRESENTS UNIQUE MOBILE PHONE USERS*	*FIGURE BASED ON ACTIVE USER ACCOUNTS, NOT UNIQUE INDIVIDUALS*

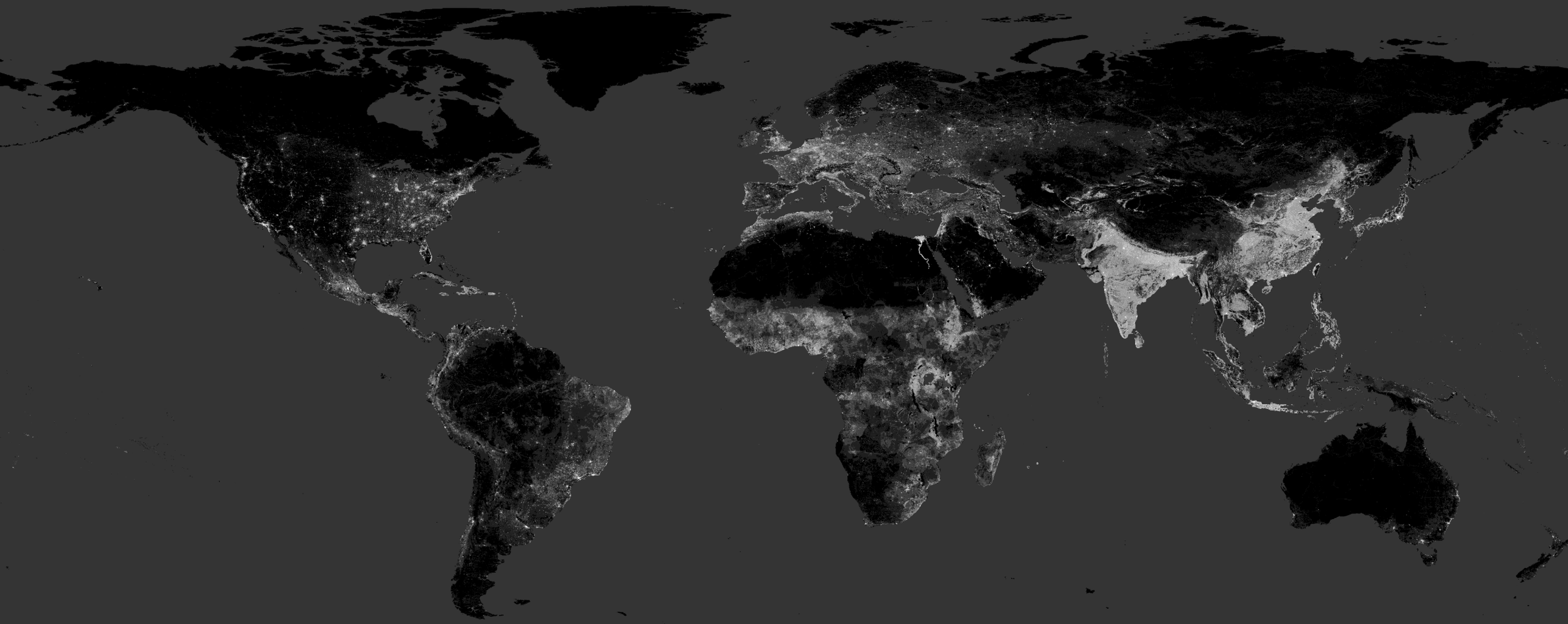
we are social

Sources: Population: UN, US Census Bureau; Internet: ITU, InternetWorldStats, CIA, national government ministries and industry associations; Social & Mobile Social: Facebook, Tencent, VKontakte, LiveInternet.ru, Nikkei, VentureBeat, Niki Aghaei; Mobile: GSMA Intelligence.

World Coverage



World Population



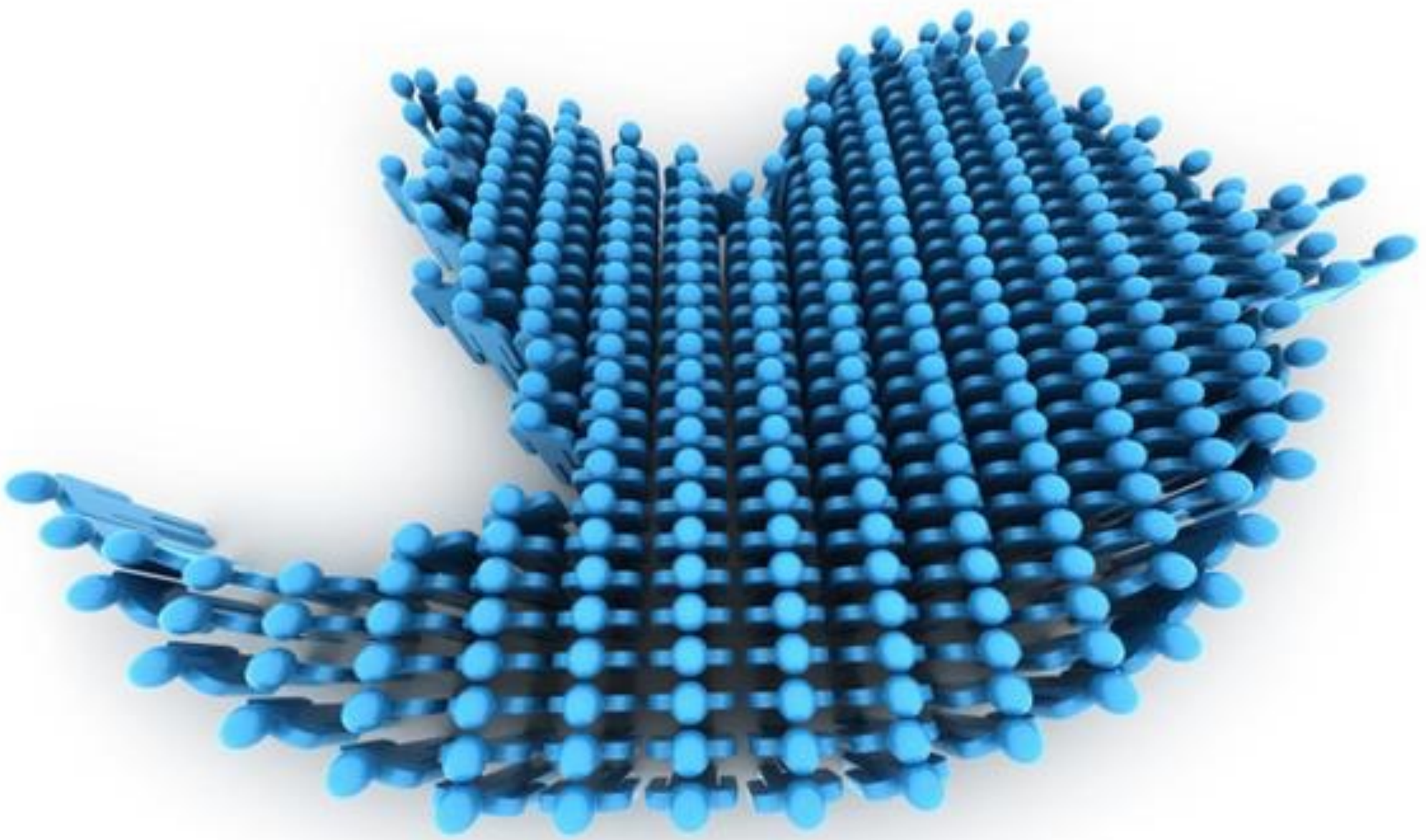
@bgoncalves

www.bgoncalves.com

Geolocated Activity

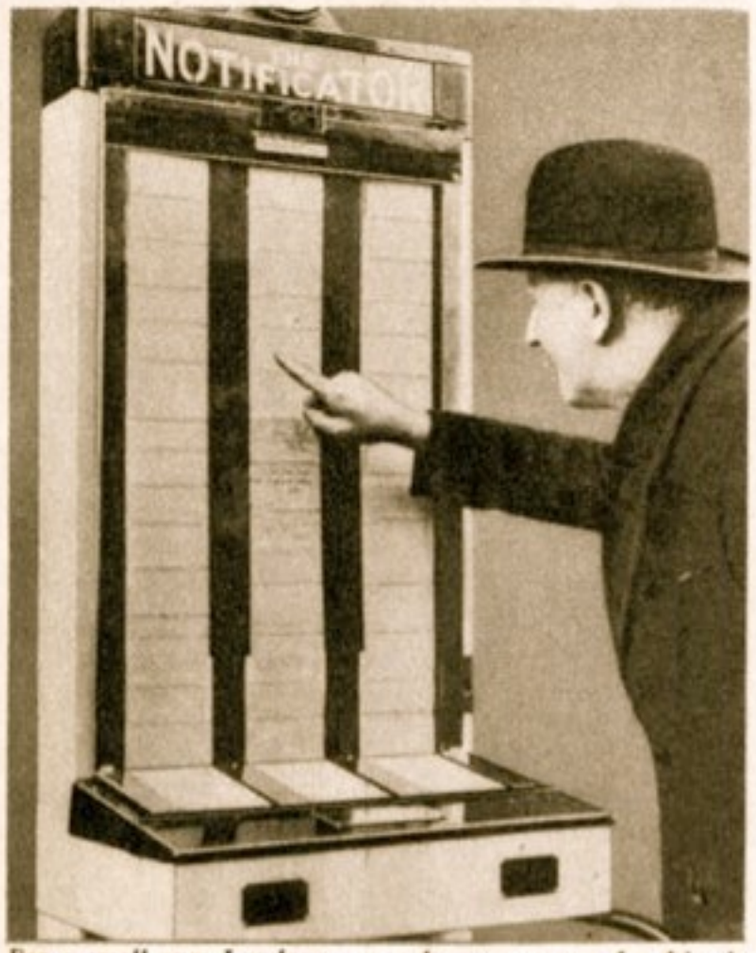
PLoS One 8, E61981 (2013)

Twitter



Twitter

Robot Messenger Displays Person-to-Person Notes In Public



For a small sum Londoners may leave messages for friends in public places. When written on "notificator," message moves up behind window, remaining in view for two hours.

TO AID persons who wish to make or cancel appointments or inform friends of their whereabouts, a robot message carrier has been introduced in London, England.

Known as the "notificator," the new machine is installed in streets, stores, railroad stations or other public places where individuals may leave messages for friends.

The user walks up on a small platform in front of the machine, writes a brief message on a continuous strip of paper and drops a coin in the slot. The inscription moves up behind a glass panel where it remains in public view for at least two hours so that the person for whom it is intended may have sufficient time to observe the note at the appointed place. The machine is similar in appearance to a candy-vending device.

Source: Modern Mechanix (Aug, 1935)

Anatomy of a Tweet

Anatomy of a Tweet


```
[u'contributors',
 u'truncated',
 u'text',
 u'in_reply_to_status_id',
 u'id',
 u'favorite_count',
 u'source',
 u'retweeted',
 u'coordinates',
 u'entities',
 u'in_reply_to_screen_name',
 u'in_reply_to_user_id',
 u'retweet_count',
 u'id_str',
 u'favorited',
 u'user',
 u'geo',
 u'in_reply_to_user_id_str',
 u'possibly_sensitive',
 u'lang',
 u'created_at',
 u'in_reply_to_status_id_str',
 u'place',
 u'metadata']
```

Anatomy of a Tweet

```
[u'contributors',
 u'truncated',
 u'text',
 u'in_reply_to_status_id',
 u'id',
 u'favorite_count',
 u'source',
 u'retweeted',
 u'coordinates',
 u'entities',
 u'in_reply_to_screen_name',
 u'in_reply_to_user_id',
 u'retweet_count',
 u'id_str',
 u'favorited',
 u'user',
 u'geo',
 u'in_reply_to_user_id_str',
 u'possibly_sensitive',
 u'lang',
 u'created_at',
 u'in_reply_to_status_id_str',
 u'place',
 u'metadata']
```

```
[u'follow_request_sent',
 u'profile_use_background_image',
 u'default_profile_image',
 u'id',
 u'profile_background_image_url_https',
 u'verified',
 u'profile_text_color',
 u'profile_image_url_https',
 u'profile_sidebar_fill_color',
 u'entities',
 u'followers_count',
 u'profile_sidebar_border_color',
 u'id_str',
 u'profile_background_color',
 u'listed_count',
 u'is_translation_enabled',
 u'utc_offset',
 u'statuses_count',
 u'description',
 u'friends_count',
 u'location',
 u'profile_link_color',
 u'profile_image_url',
 u'following',
 u'geo_enabled',
 u'profile_banner_url',
 u'profile_background_image_url',
 u'screen_name',
 u'lang',
```

```
 u'profile_background_tile',
 u'favourites_count',
 u'name',
 u'notifications',
 u'url',
 u'created_at',
 u'contributors_enabled',
 u'time_zone',
 u'protected',
 u'default_profile',
 u'is_translator']
```

Anatomy of a Tweet

```
[u'contributors',
 u'truncated',
 u'text',
 u'in_reply_to_status_id',
 u'id',
 u'favorite_count',
 u'source',
 u'retweeted',
 u'coordinates',
 u'entities',
 u'in_reply_to_screen_name',
 u'in_reply_to_user_id',
 u'retweet_count',
 u'id_str',
 u'favorited',
 u'user',
 u'geo',
 u'in_reply_to_user_id_str',
 u'possibly_sensitive',
 u'lang',
 u'created_at',
 u'in_reply_to_status_id_str',
 u'place',
 u'metadata']
```

```
u"I'm at Terminal Rodovi\xe1rio de Feira de Santana
 (Feira de Santana, BA) http://t.co/WirvdHwYMq"
```

```
u'<a href="http://foursquare.com" rel="nofollow">
  foursquare</a>'
```

```
[u'symbols',
 u'user_mentions',
 u'hashtags',
 u'urls']
```

```
[u'type',
 u'coordinates']
```

Anatomy of a Tweet

```
[u'contributors',
 u'truncated',
 u'text',
 u'in_reply_to_status_id',
 u'id',
 u'favorite_count',
 u'source',
 u'retweeted',
 u'coordinates',
 u'entities',
 u'in_reply_to_screen_name',
 u'in_reply_to_user_id',
 u'retweet_count',
 u'id_str',
 u'favorited',
 u'user',
 u'geo',
 u'in_reply_to_user_id_str',
 u'possibly_sensitive',
 u'lang',
 u'created_at',
 u'in_reply_to_status_id_str',
 u'place',
 u'metadata']
```

```
u"I'm at Terminal Rodovi\xe1rio de Feira de Santana
 (Feira de Santana, BA) http://t.co/WirvdHwYMq"
```

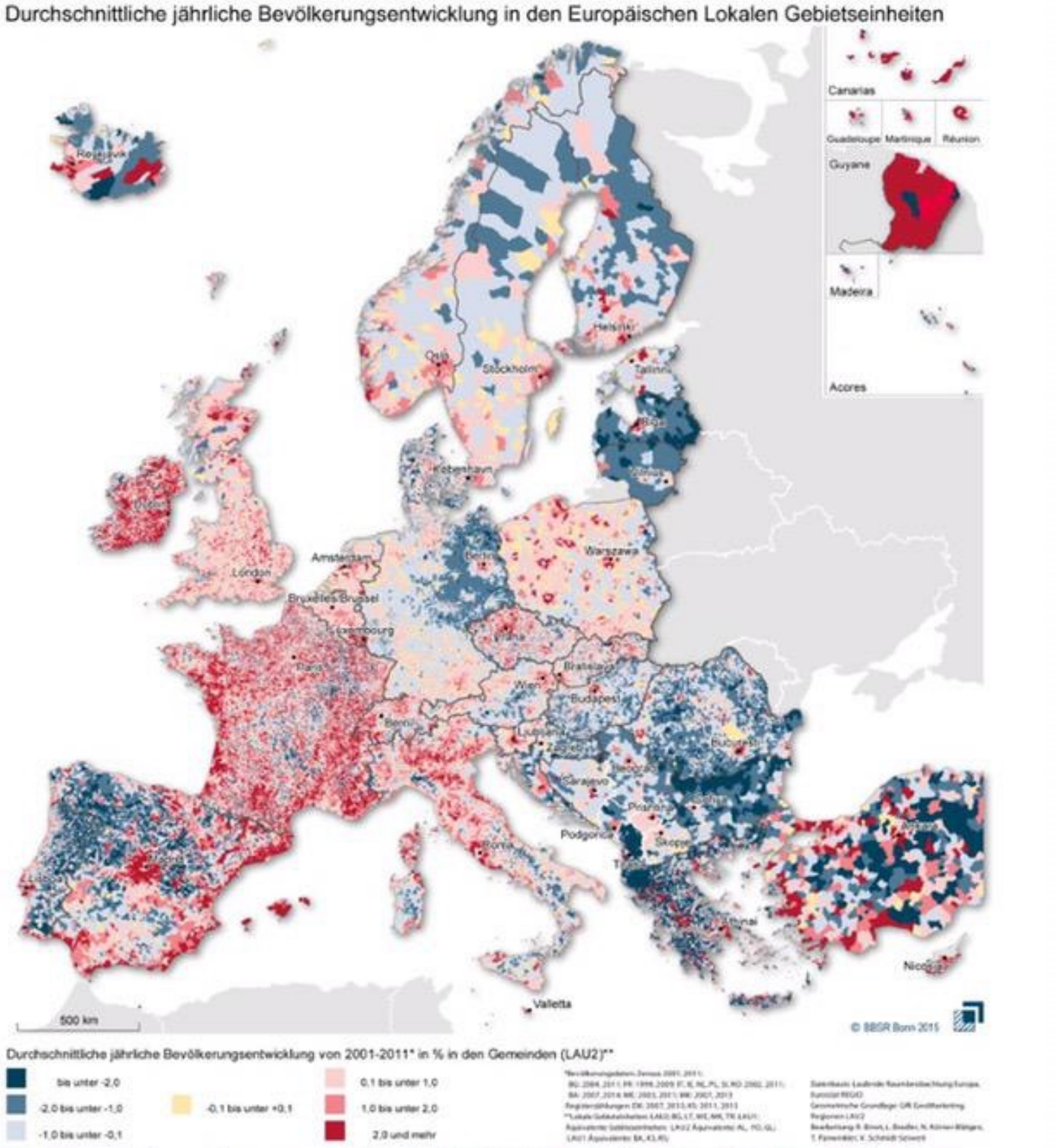
```
u'<a href="http://foursquare.com" rel="nofollow">
  foursquare</a>'
```

```
[u'symbols',
 u'user_mentions',
 u'hashtags',
 u'urls']
```

```
{u'display_url': u'4sq.com/1k5MeYF',
 u'expanded_url': u'http://4sq.com/1k5MeYF',
 u'indices': [70, 92],
 u'url': u'http://t.co/WirvdHwYMq'}
```

```
[u'type',
 u'coordinates']
```

Biases



Smartphone Ownership Highest Among Young Adults, Those With High Income/Education Levels

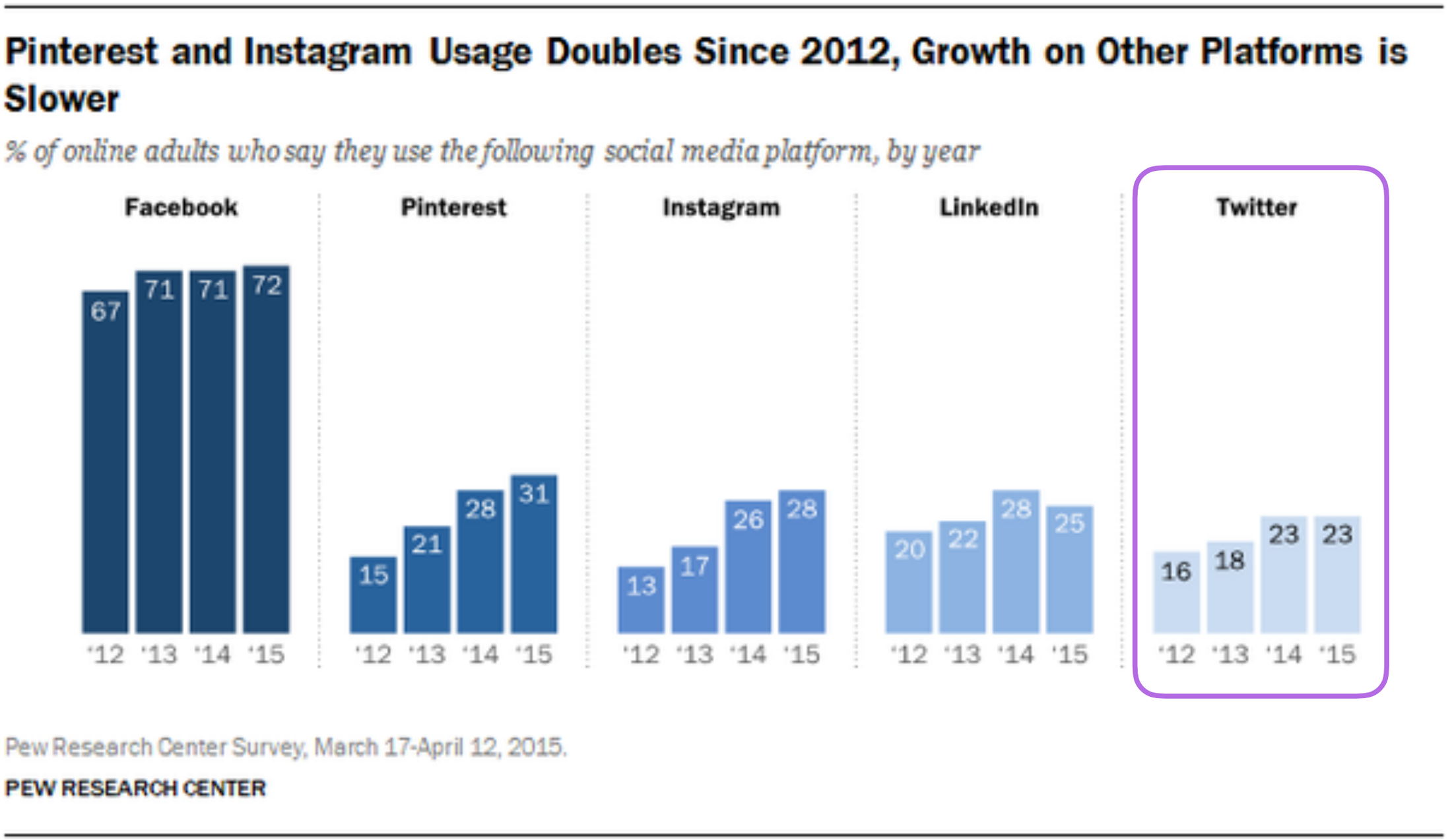
% of U.S. adults in each group who own a smartphone

All adults	**64%**
Male	66
Female	63
18-29	85
30-49	79
50-64	54
65+	27
White, non-Hispanic	61
Black, non-Hispanic	70
Hispanic	71
HS grad or less	52
Some college	69
College+	78
Less than $30,000/yr	50
$30,000-$49,999	71
$50,000-$74,999	72
$75,000 or more	84
Urban	68
Suburban	66
Rural	52

Combined analysis of Pew Research Center surveys conducted Dec. 4-7 and Dec. 18-21, 2014.

PEW RESEARCH CENTER

Biases


Pinterest and Instagram Usage Doubles Since 2012, Growth on Other Platforms is Slower
% of online adults who say they use the following social media platform, by year
Facebook
67
71
71
72
'12 '13 '14 '15
Pinterest
15
21
28
31
'12 '13 '14 '15
Instagram
13
17
26
28
'12 '13 '14 '15
LinkedIn
20
22
28
25
'12 '13 '14 '15
Twitter
16
18
23
23
'12 '13 '14 '15
Pew Research Center Survey, March 17-April 12, 2015.
PEW RESEARCH CENTER

Age Distribution

PLoS One 10, e0115545 (2015)

Twitter users

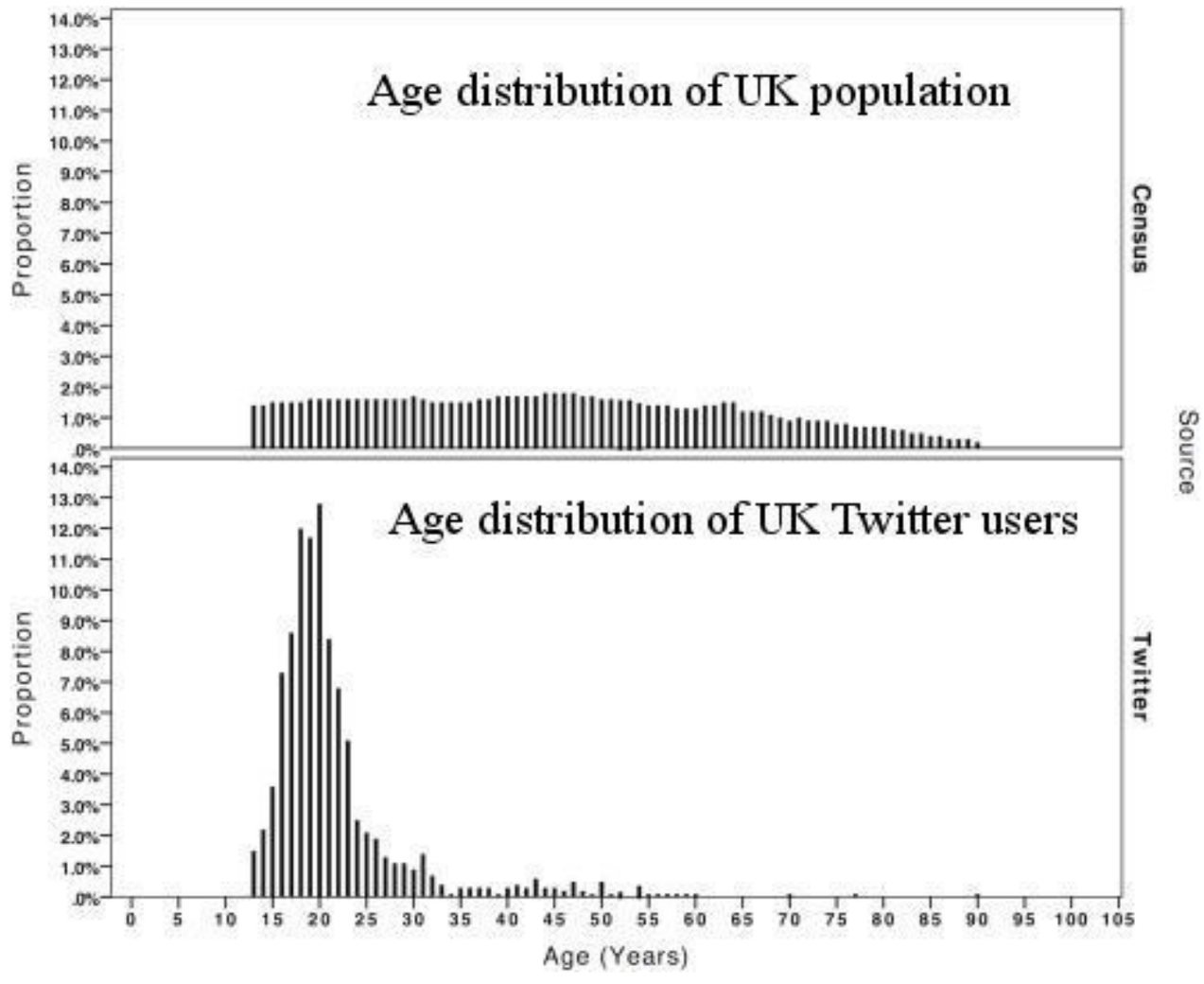
Among online adults, the % who use Twitter

	2013	2014
All internet users	*18%*	*23%**
Men	17	24*
Women	18	21
White, Non-Hispanic	16	21 *
Black, Non-Hispanic	29	27
Hispanic	16	25
18-29	31	37
30-49	19	25
50-64	9	12
65+	5	10*
High school grad or less	17	16
Some college	18	24
College+ (n= 685)	18	30*
Less than $30,000/yr	17	20
$30,000-$49,999	18	21
$50,000-$74,999	15	27*
$75,000+	19	27*
Urban	18	25*
Suburban	19	23
Rural	11	17

Source: Pew Research Center's Internet Project September Combined Omnibus Survey, September 11-14 & September 18-21, 2014. N=1,597 internet users ages 18+. The margin of error for all internet users is +/- 2.9 percentage points. 2013 data from Pew Internet August Tracking Survey, August 07 – September 16, 2013, n= 1,445 internet users ages 18+.

Note: Percentages marked with an asterisk (*) represent a significant change from 2013. Results are significant at the 95% confidence level using an independent z-test.

PEW RESEARCH CENTER



Age Distribution

Twitter users

Among online adults, the % who use Twitter

	2013	2014
All internet users	*18%*	*23%**
Men	17	24*
Women	18	21
White, Non-Hispanic	16	21 *
Black, Non-Hispanic	29	27
Hispanic	16	25
18-29	31	37
30-49	19	25
50-64	9	12
65+	5	10*
High school grad or less	17	16
Some college	18	24
College+ (n= 685)	18	30*
Less than $30,000/yr	17	20
$30,000-$49,999	18	21
$50,000-$74,999	15	27*
$75,000+	19	27*
Urban	18	25*
Suburban	19	23
Rural	11	17

Source: Pew Research Center's Internet Project September Combined Omnibus Survey, September 11-14 & September 18-21, 2014. N=1,597 internet users ages 18+. The margin of error for all internet users is +/- 2.9 percentage points. 2013 data from Pew Internet August Tracking Survey, August 07 – September 16, 2013, n= 1,445 internet users ages 18+.

Note: Percentages marked with an asterisk (*) represent a significant change from 2013. Results are significant at the 95% confidence level using an independent z-test.

PEW RESEARCH CENTER

Facebook Demographics

Among internet users, the % who use Facebook

	Internet users
Total	*72%*
Men	66
Women	77
White, Non-Hispanic	70
Black, Non-Hispanic (n=85)	67
Hispanic	75
18-29	82
30-49	79
50-64	64
65+	48
High school grad or less	71
Some college	72
College+	72
Less than $30,000/yr	73
$30,000-$49,999	72
$50,000-$74,999	66
$75,000+	78
Urban	74
Suburban	72
Rural	67

Source: Pew Research Center, March 17-April 12, 2015.

PEW RESEARCH CENTER

Instagram Demographics

Among internet users, the % who use Instagram

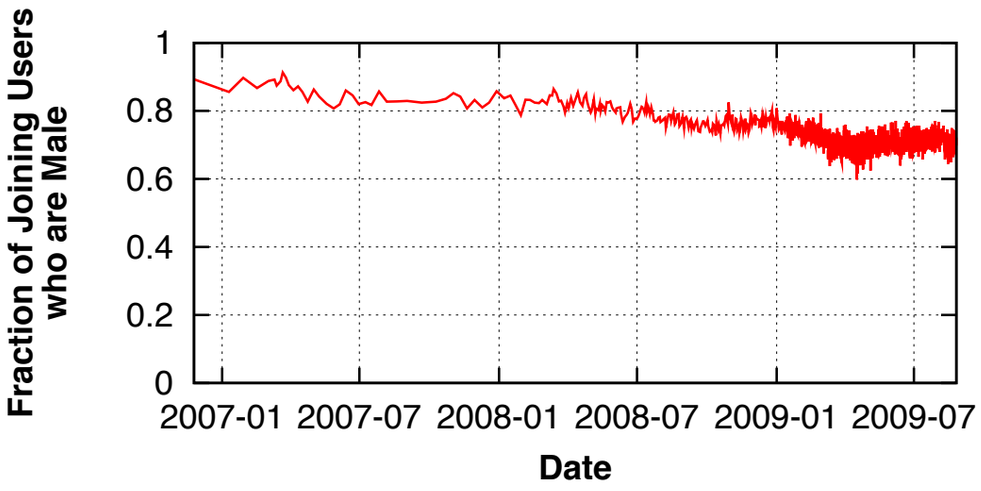
	Internet user
Total	*28%*
Men	24
Women	31
White, Non-Hispanic	21
Black, Non-Hispanic (n=85)	47
Hispanic	38
18-29	55
30-49	28
50-64	11
65+	4
High school grad or less	25
Some college	32
College+	26
Less than $30,000/yr	26
$30,000-$49,999	27
$50,000-$74,999	30
$75,000+	26
Urban	32
Suburban	28
Rural	18

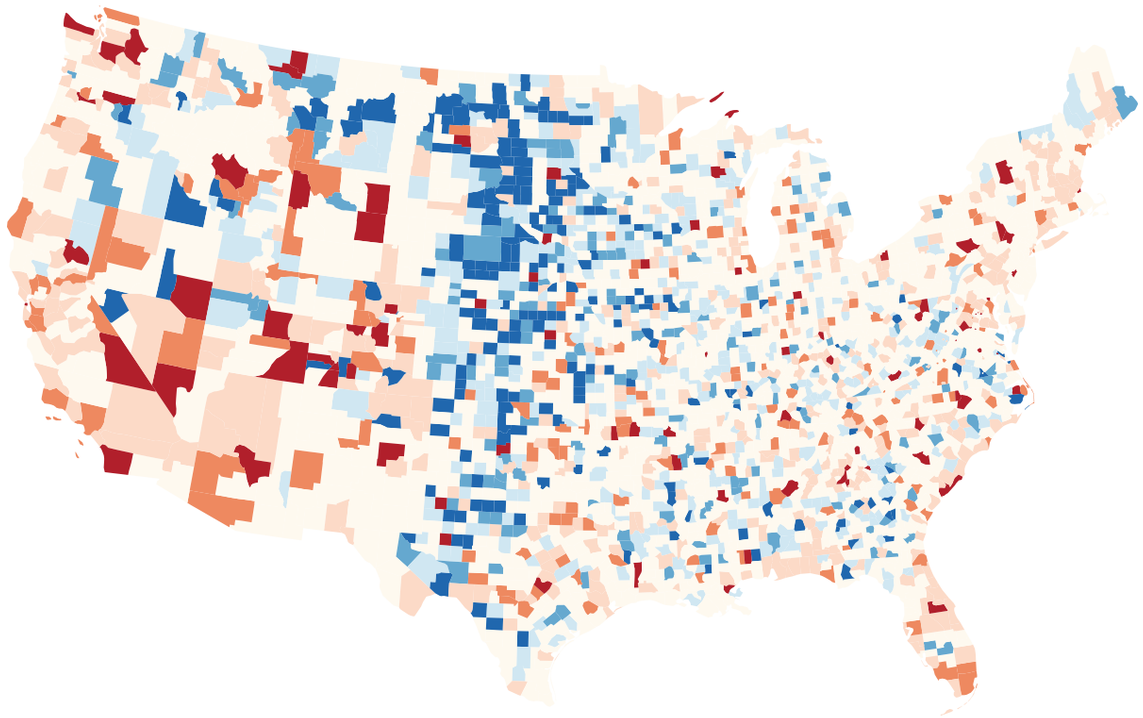
Source: Pew Research Center, March 17-April 12, 2015.

PEW RESEARCH CENTER

Demographics

ICWSM'11, 375 (2011)





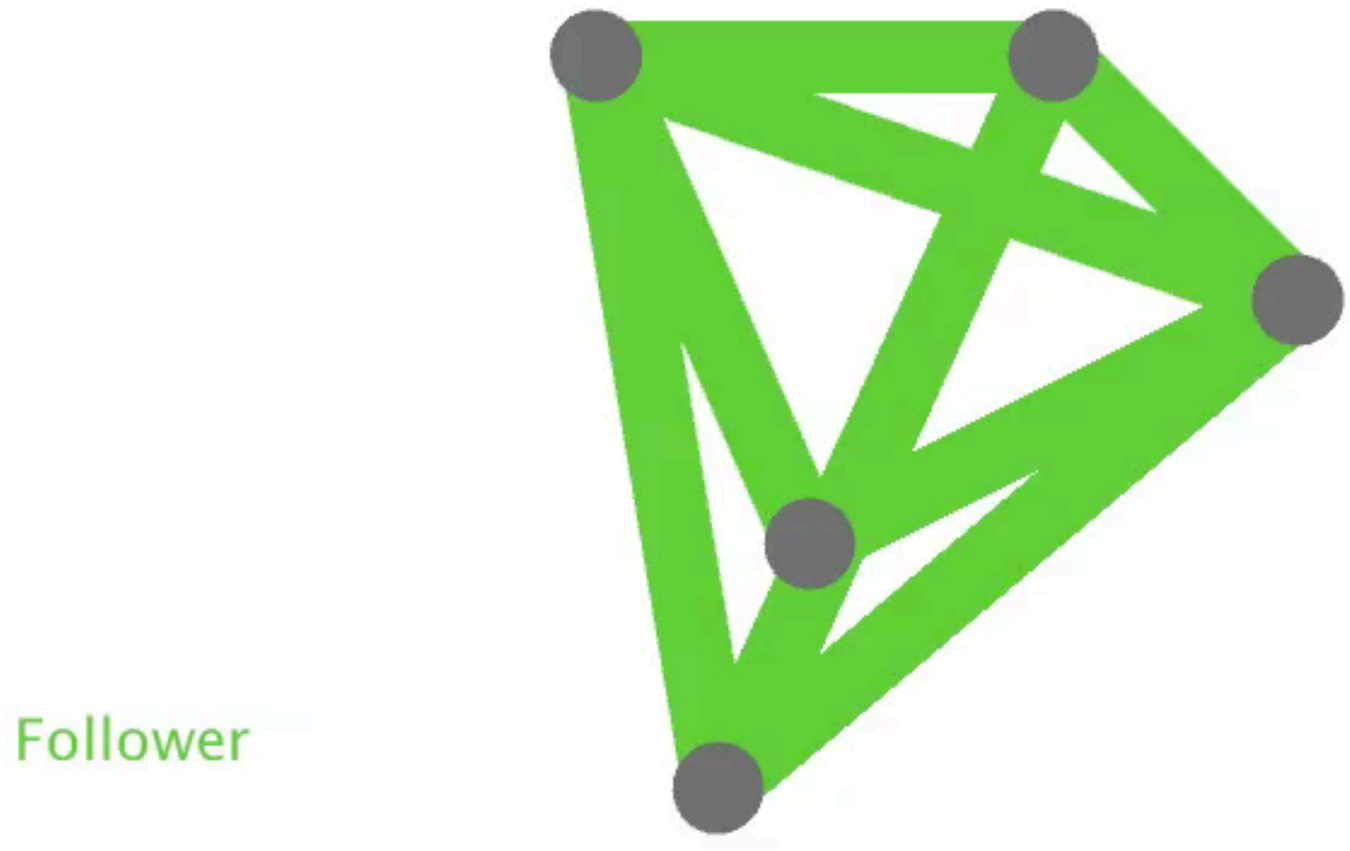
(a) Caucasian (non-hispanic)

(b) African-American

(c) Asian or Pacific Islander

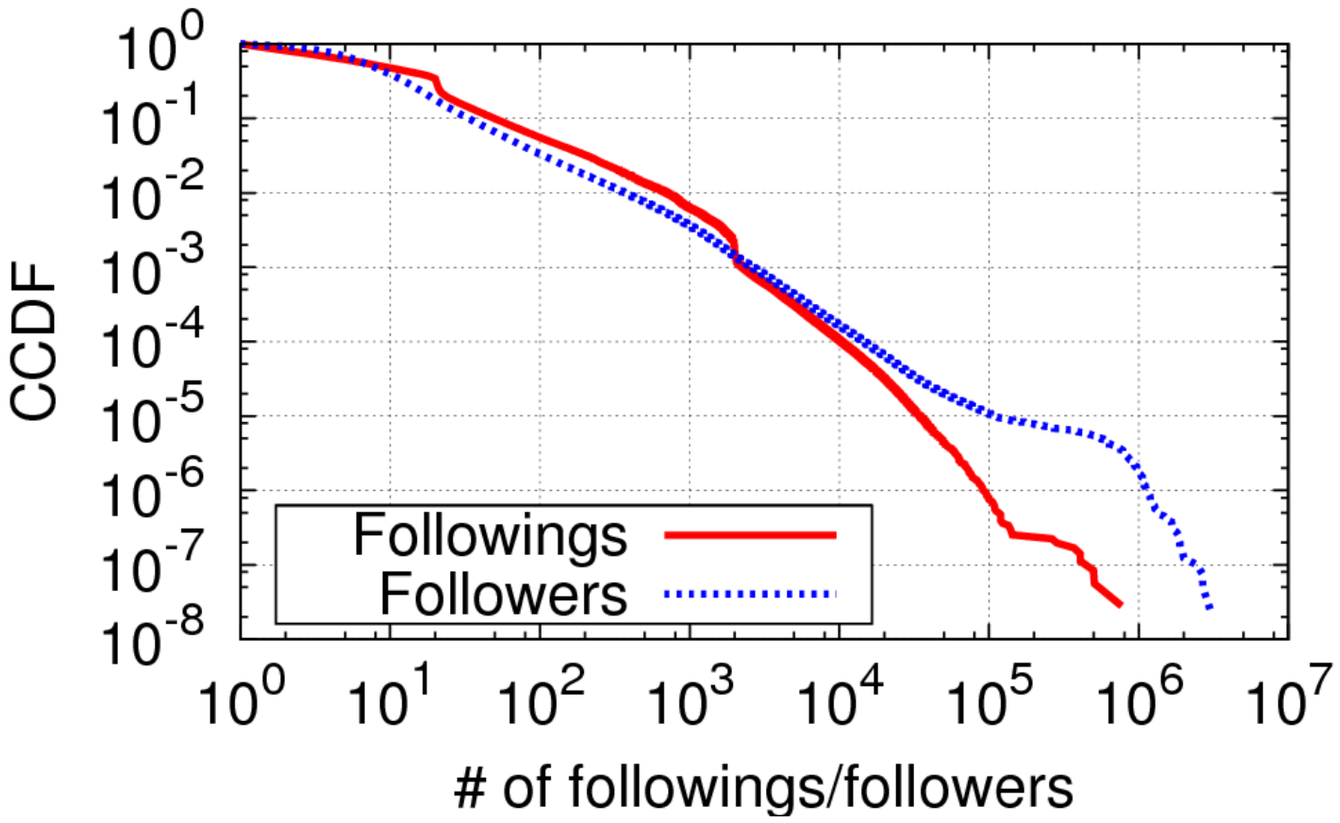
(d) Hispanic

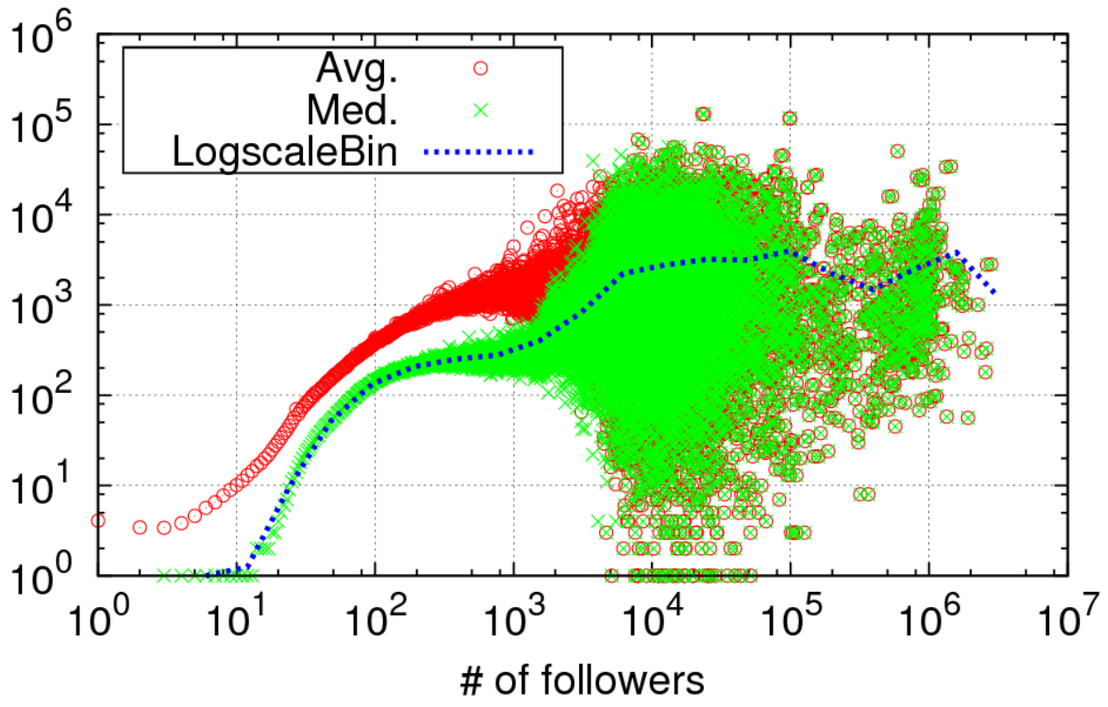
Multilayer Network

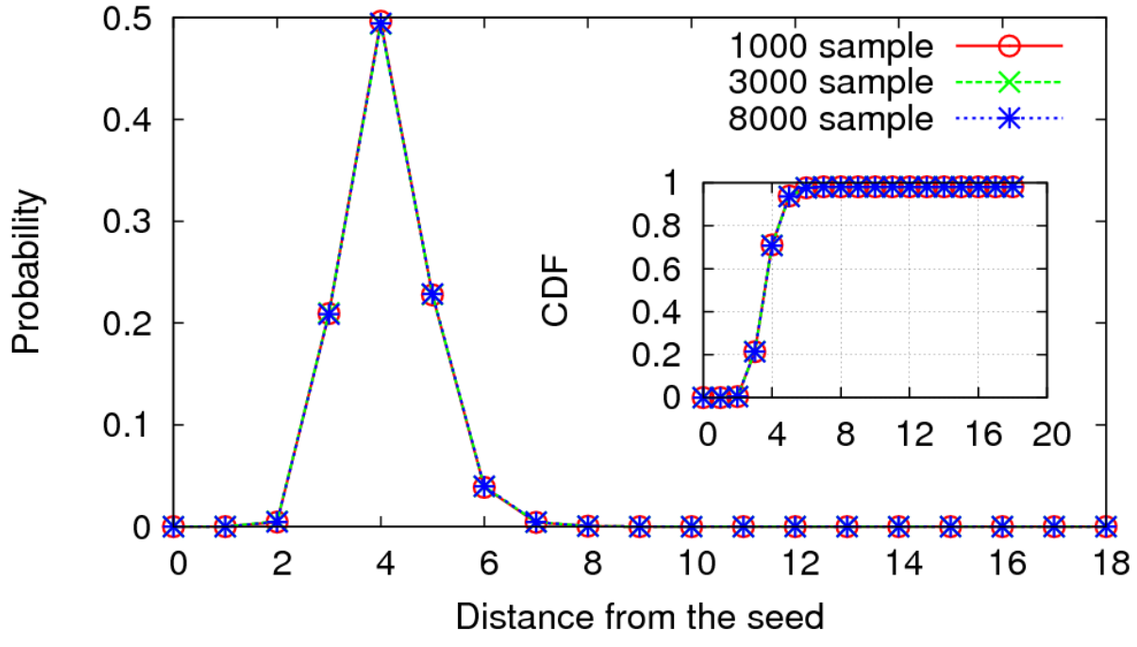


Twitter Network

WWW'10, 591 (2010)

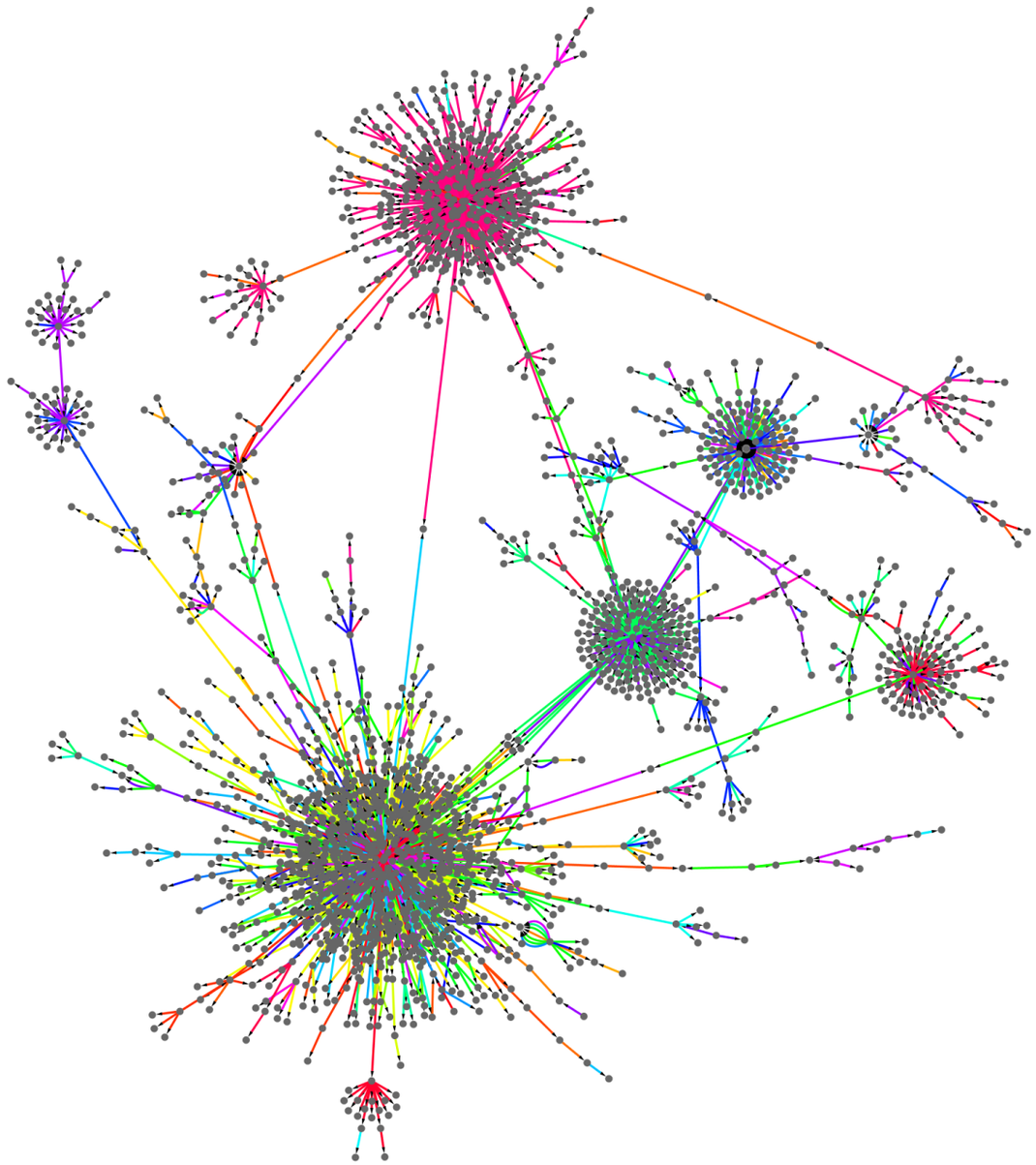


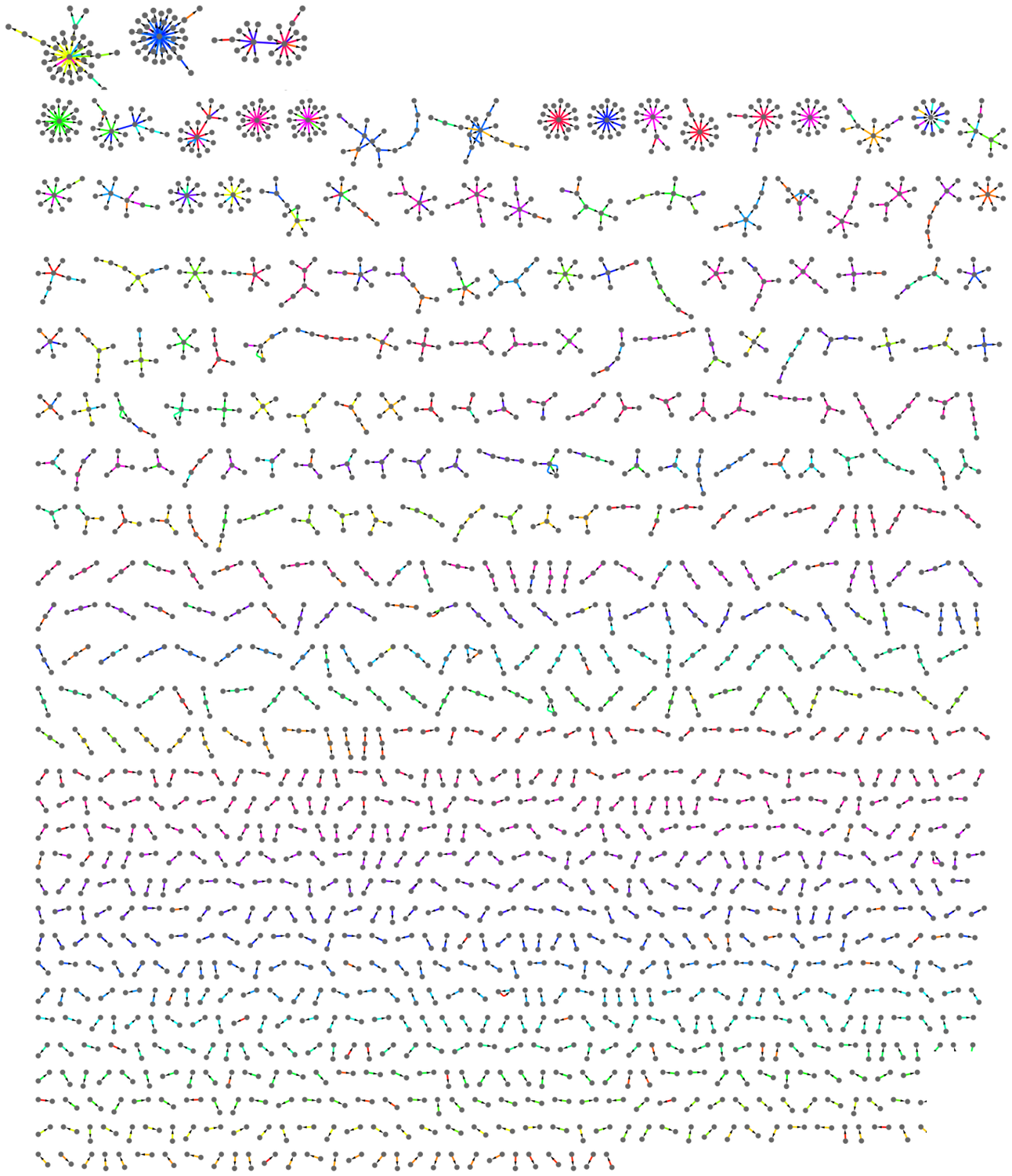




Retweet Trees

WWW'10, 591 (2010)

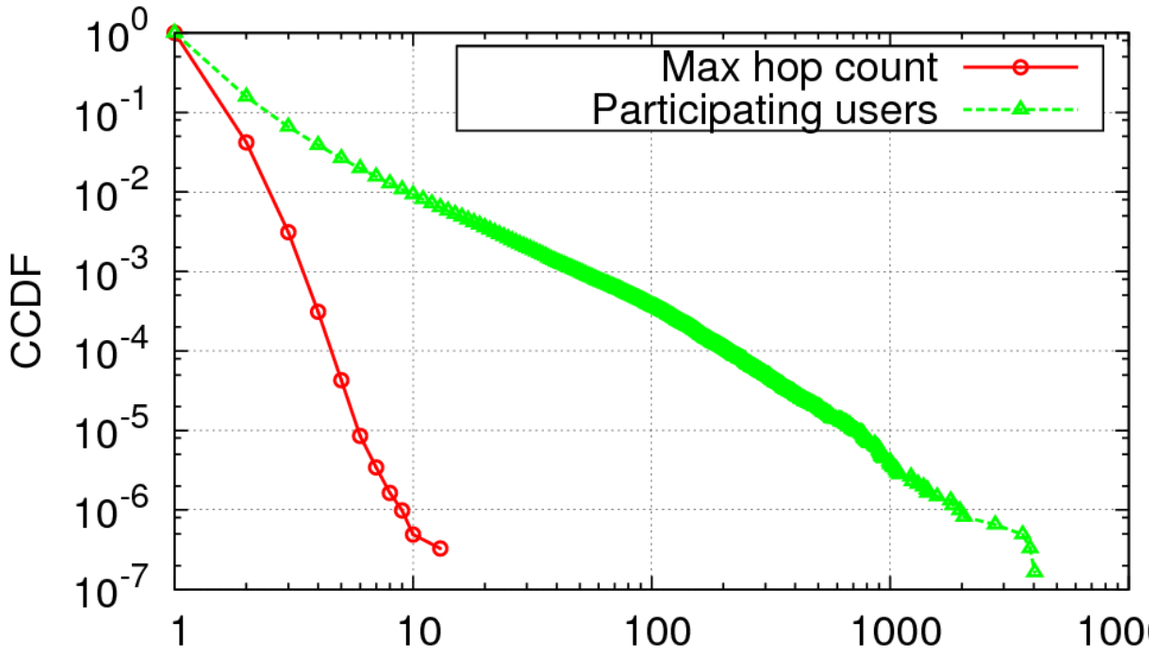




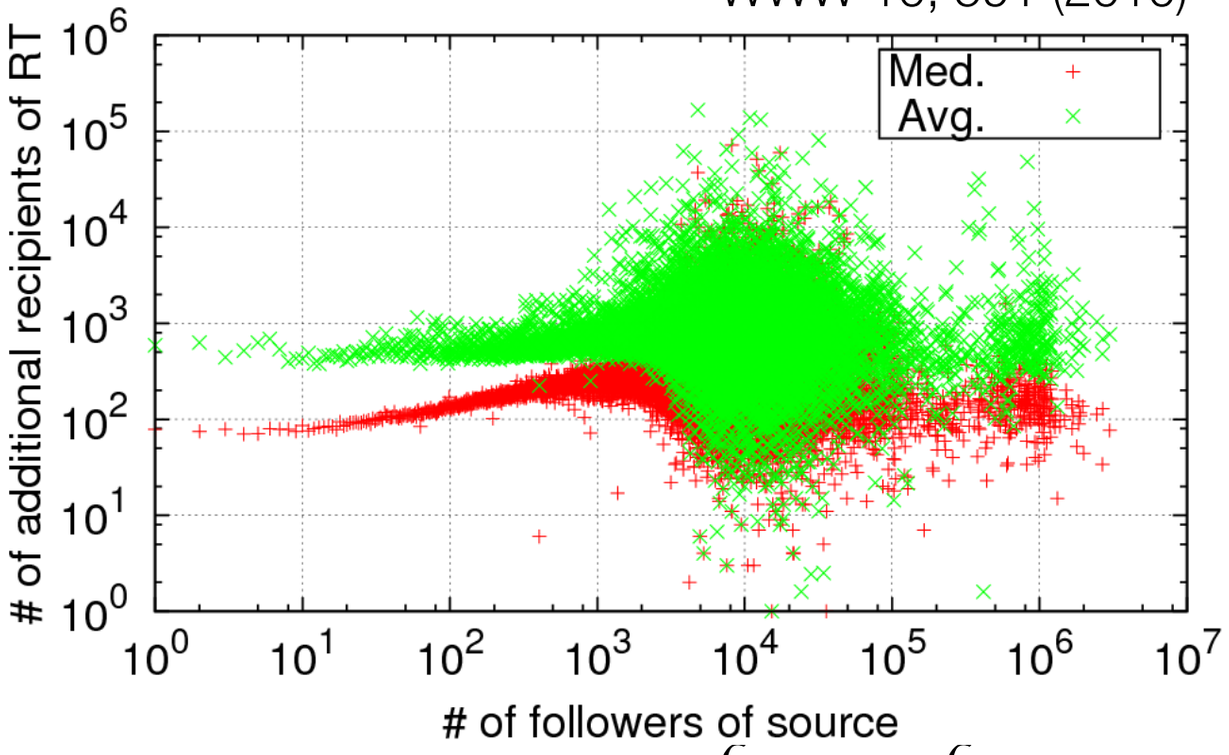
@bgoncalves

Retweets Trees

WWW'10, 591 (2010)

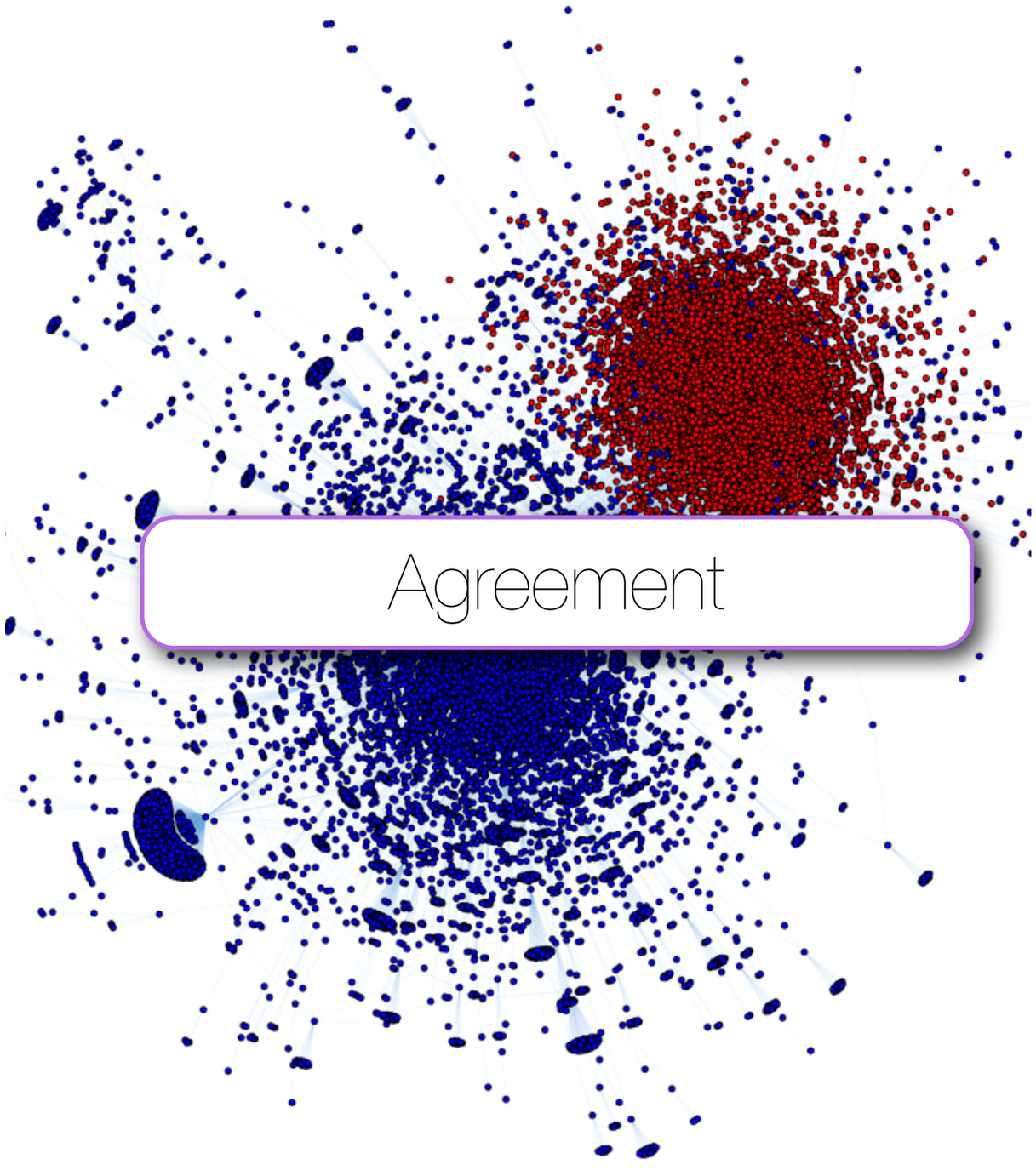


WWW'10, 591 (2010)

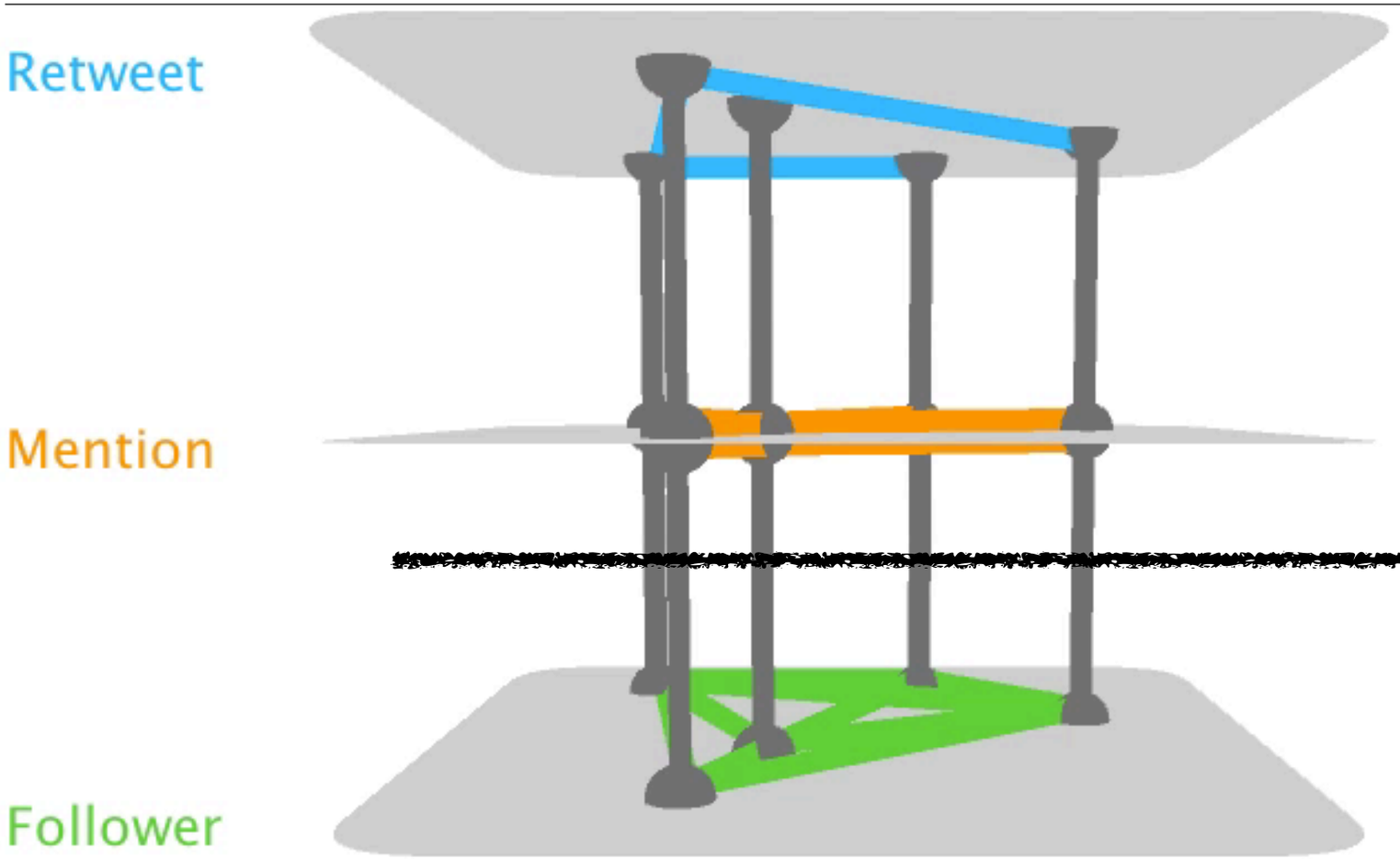


Link Function

ICWSM'11, 89 (2011)



Multilayer Network



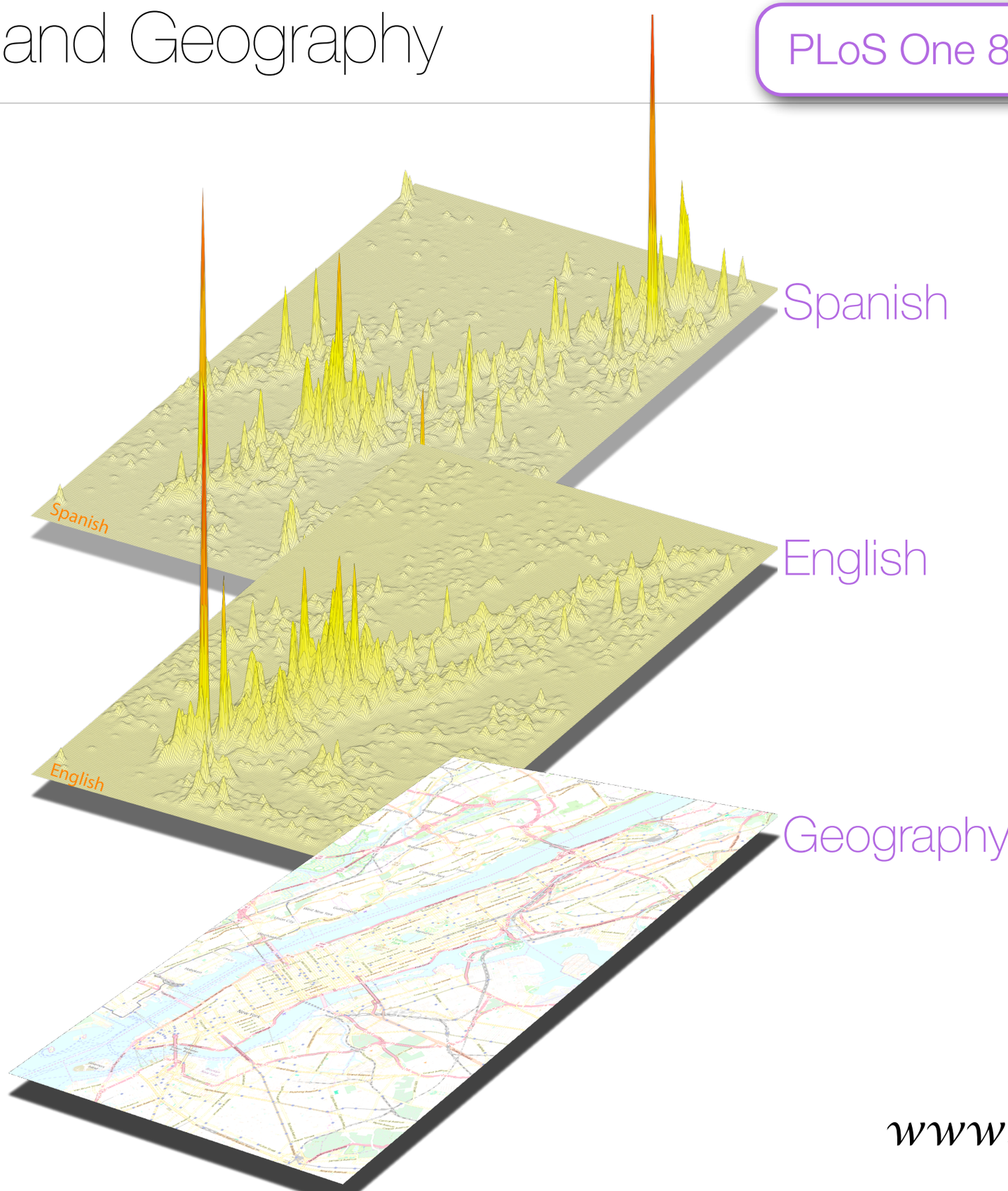
Information Layer(s)

Social Layer(s)

Geographical Layer(s)

Language and Geography

PLoS One 8, E61981 (2013)



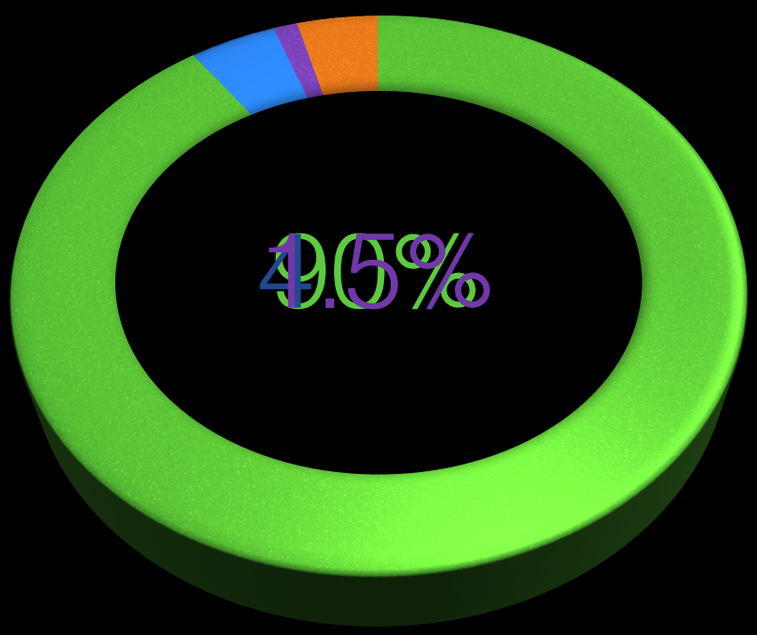
@bgoncalves

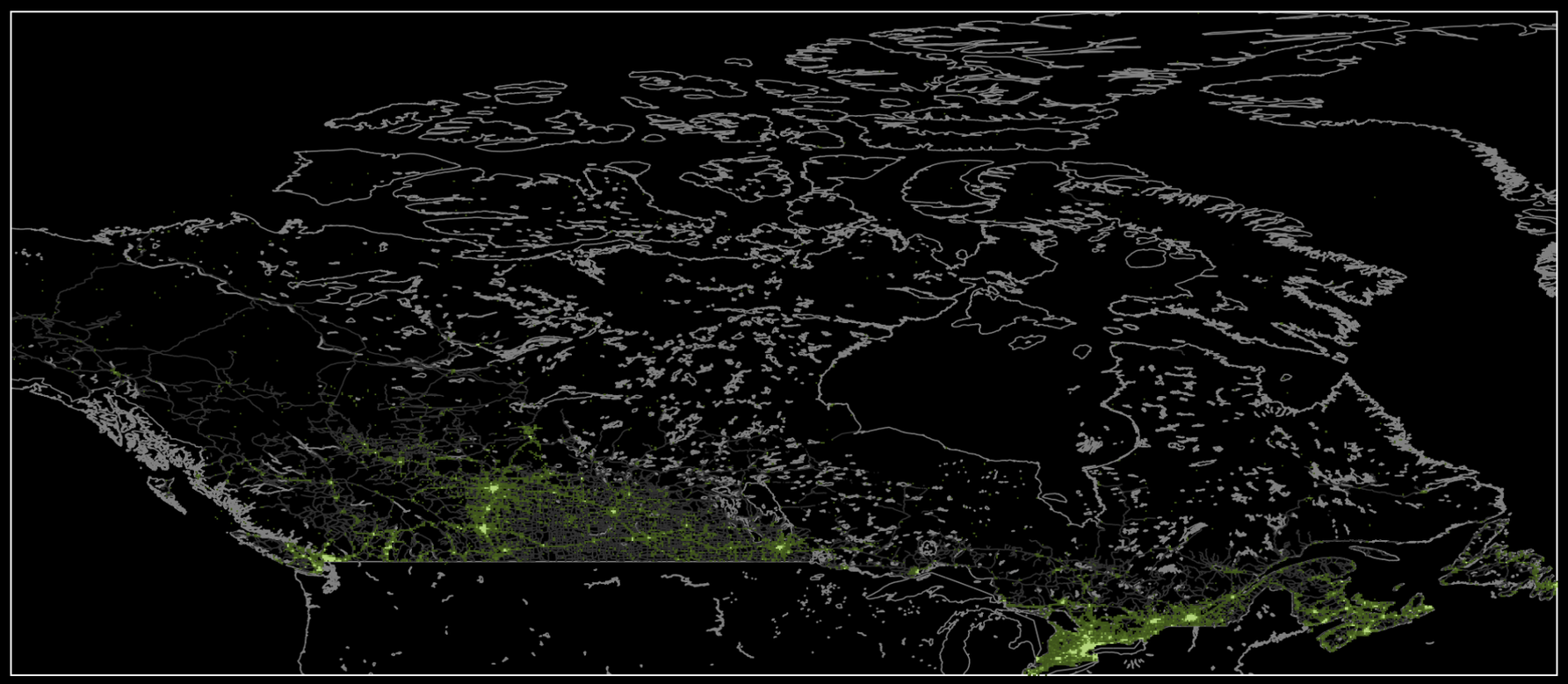
www.bgoncalves.com

Signal By Language

PLoS One 8, E61981 (2013)

- English
- French
- Spanish
- Other

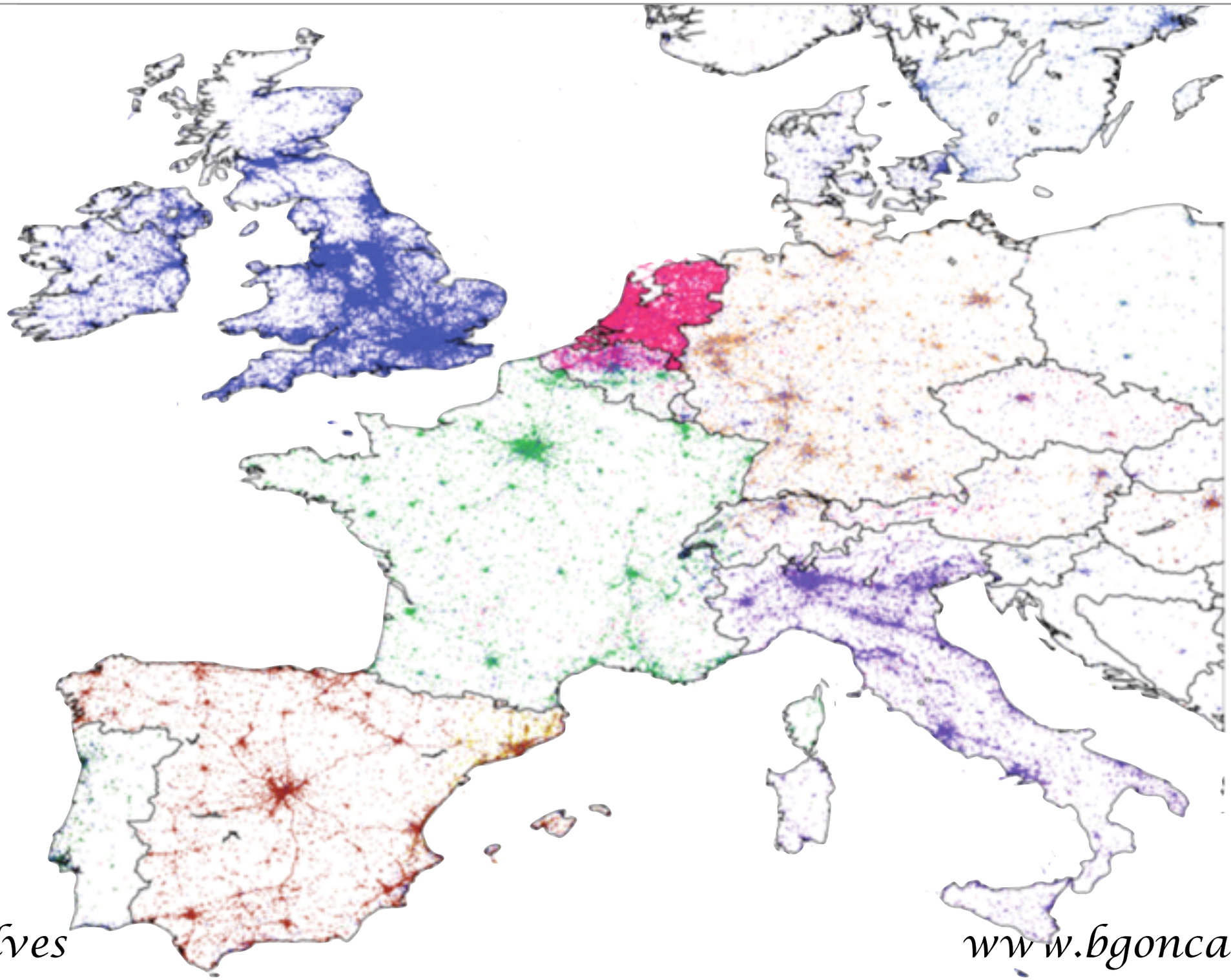




@bgoncalves

Language Geography

PLoS One 8, E61981 (2013)

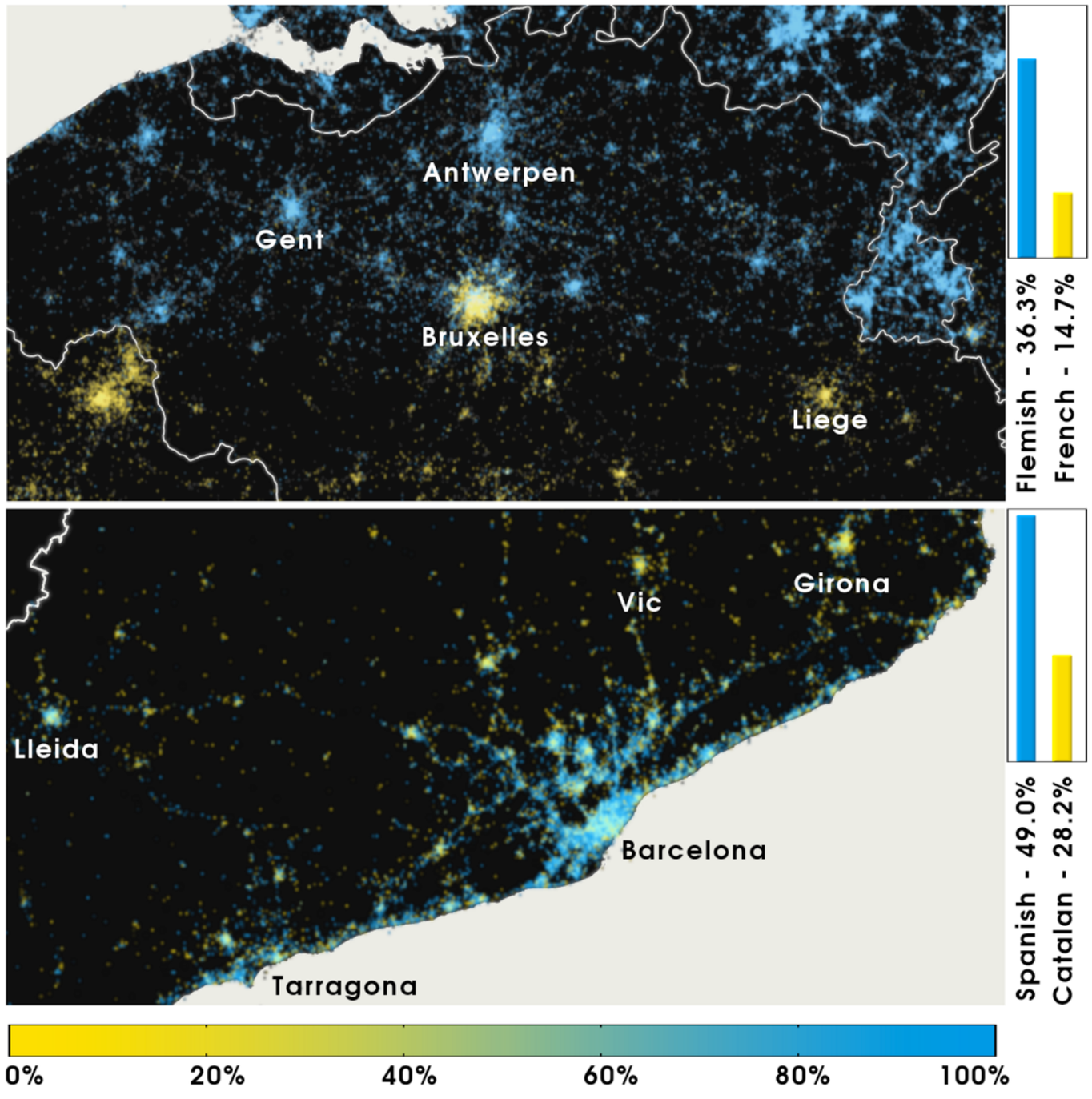


@bgoncalves

www.bgoncalves.com

Language Distribution

PLoS One 8, E61981 (2013)



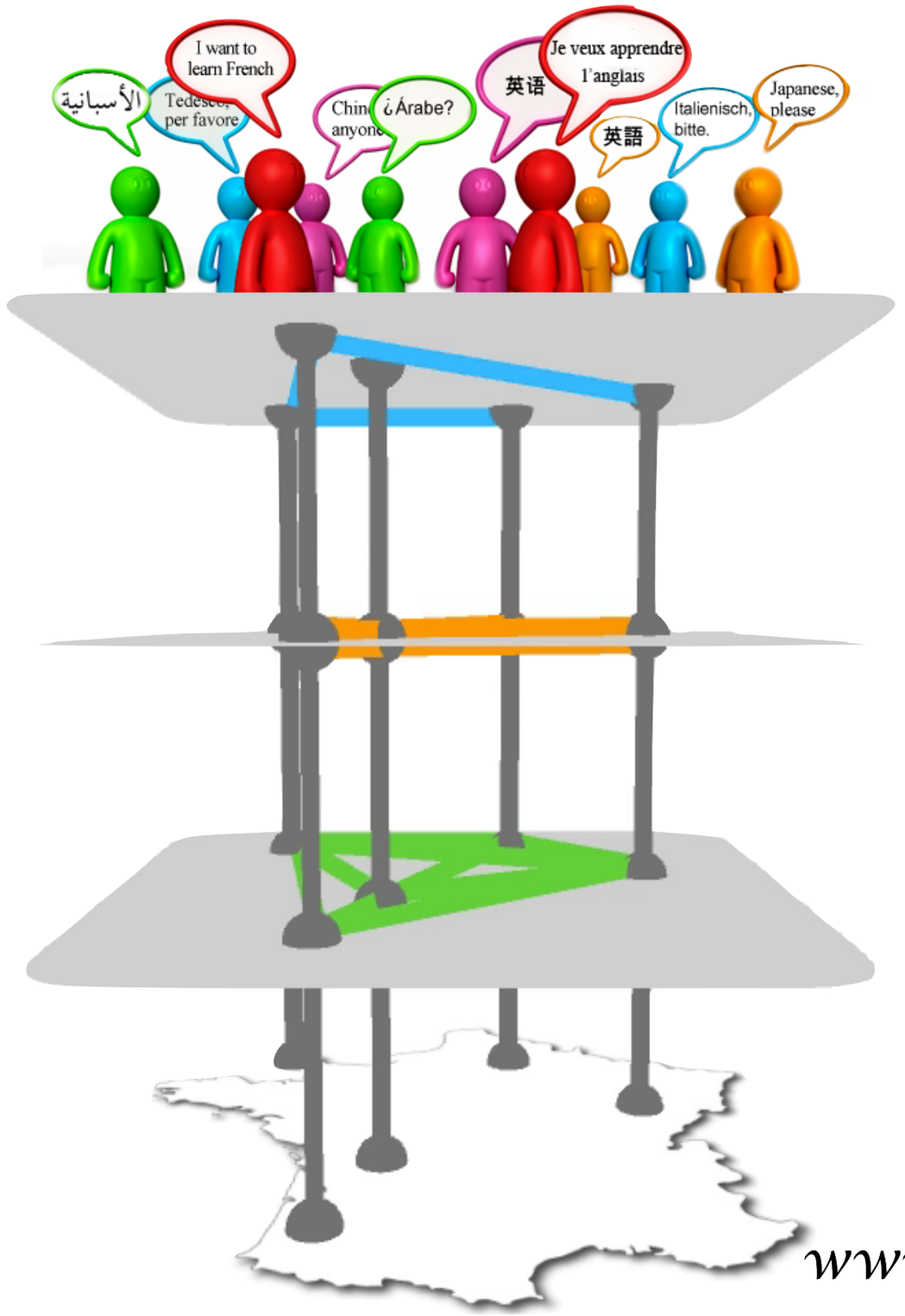
Multilayer Network

Language

Retweet

Mention

Follower



Foursquare


foursquare
CHECK-IN
HERE
Check in to unlock specials, meet up with
friends and explore what's nearby.
foursquare.com
@foursquare

Foursquare


12:29
Four Square for Dummies
Sponso
FAQ
Videos
Twitter
Suppor
foursquare
I'm looking for...
About Foursquare
What is Foursquare?
Foursquare is a free app that helps you and your friends make the most of where you are. When you're out and about, use Foursquare to share and save the places you visit. And, when you're looking for inspiration for what to do next, we'll give you personalized recommendations and deals based on where you, your friends, and people with your tastes have been.
Whether you're setting off on a trip around the world, coordinating a night out with friends, or trying to pick out the best dish at your local restaurant, Foursquare is the perfect companion.
Fetched

Anatomy of a Checkin

Anatomy of a Checkin

```
[u'venue',
 u'like',
 u'photos',
 u'source',
 u'visibility',
 u'entities',
 u'shout',
 u'timeZoneOffset',
 u'type',
 u'id',
 u'createdAt',
 u'likes']
```

Anatomy of a Checkin

```
[u'venue',
 u'like',
 u'photos',
 u'source',
 u'visibility',
 u'entities',
 u'shout',
 u'timeZoneOffset',
 u'type',
 u'id',
 u'createdAt',
 u'likes']
```

```
[u'verified',
 u'name',
 u'url',
 u'like',
 u'contact',
 u'location',
 u'stats',
 u'id',
 u'categories',
 u'likes']
```

```
[{u'indices': [112, 137],
  u'object': {u'url': u'http://go.nasa.gov/9kpN5g'},
  u'type': u'url'}]
```

```
u"We've got a new 'Curiosity Explorer' badge. Explore your
curiosity at science museums & planetariums to earn it http://
go.nasa.gov/9kpN5g"
```

```
{u'count': 1837,
 u'groups': [{u'count': 1837, u'items': [], u'type': u'others'}],
 u'summary': u'1837 likes'}
```

Objects

- Two main types of objects:
    - Users
    - Venues
- Multiple possible actions (by Users on Venues)
    - Checkin
    - Like
    - Tip

API Limitations

- Users have privacy concerns with respect to publicly sharing their location.
    - "Stalker apps"
    - "Please Rob Me"
- Privacy is a big concern for Foursquare
    - API structure reflects this
    - "Easy" to get information on users and on venues. Connecting users to venues much harder to do.

Objects

https://developer.foursquare.com/overview/venues.html

Venues

- Venues correspond to physical locations
- Are perhaps the most important object in the Foursquare universe
- API is particularly generous, allowing for 5000 requests per hour.

Objects

https://developer.foursquare.com/docs/users/users

Venues

- Venues correspond to physical locations
- Are perhaps the most important object in the Foursquare universe
- API is particularly generous, allowing for 5000 requests per hour.

Users

- Users interact with venues in multiple ways:
    - checkin
    - leaving a "tip"
    - "liking"
- Users connect to each other through friendship/colocation

Objects

https://developer.foursquare.com/docs/checkins/checkins.html

Checkins

- Checkins are the *Raison d'être* of Foursquare.
- They connect Users with Venues providing valuable temporal and demographic information.
- Users have the option of sharing their checkins through social media
- The status text is shared along with the URL of the web version of the checkin.
- Currently, particularly hard to obtain

Mobility and Social Structure

Airline Flights

US Flight dataset

https://www.transtats.bts.gov/databases.asp?Mode_ID=1&Mode_Desc=Aviation&Subject_ID2=0

RITA | BTS | Transtats

Bruno

Secure https://www.transtats.bts.gov/databases.asp?Mode_ID=1&Mode_Desc=Aviation&Subject_ID2=0

United States Department of Transportation

Ask a Research Question | A-Z Index

Bureau of Transportation Statistics

Explore Topics and Geography | Browse Statistical Products and Data | Learn About BTS and Our Work | Newsroom

OST-R > BTS

TranStats

Search this site:

Go

Advanced Search

Resources

- Database Directory
- Glossary
- Upcoming Releases
- Data Release History

Data Finder

By Mode

- Aviation
- Maritime
- Highway
- Transit
- Rail
- Pipeline
- Bike/Pedestrian
- Other

By Subject

- Safety
- Freight Transport
- Passenger Travel
- Infrastructure

Data Library: Aviation

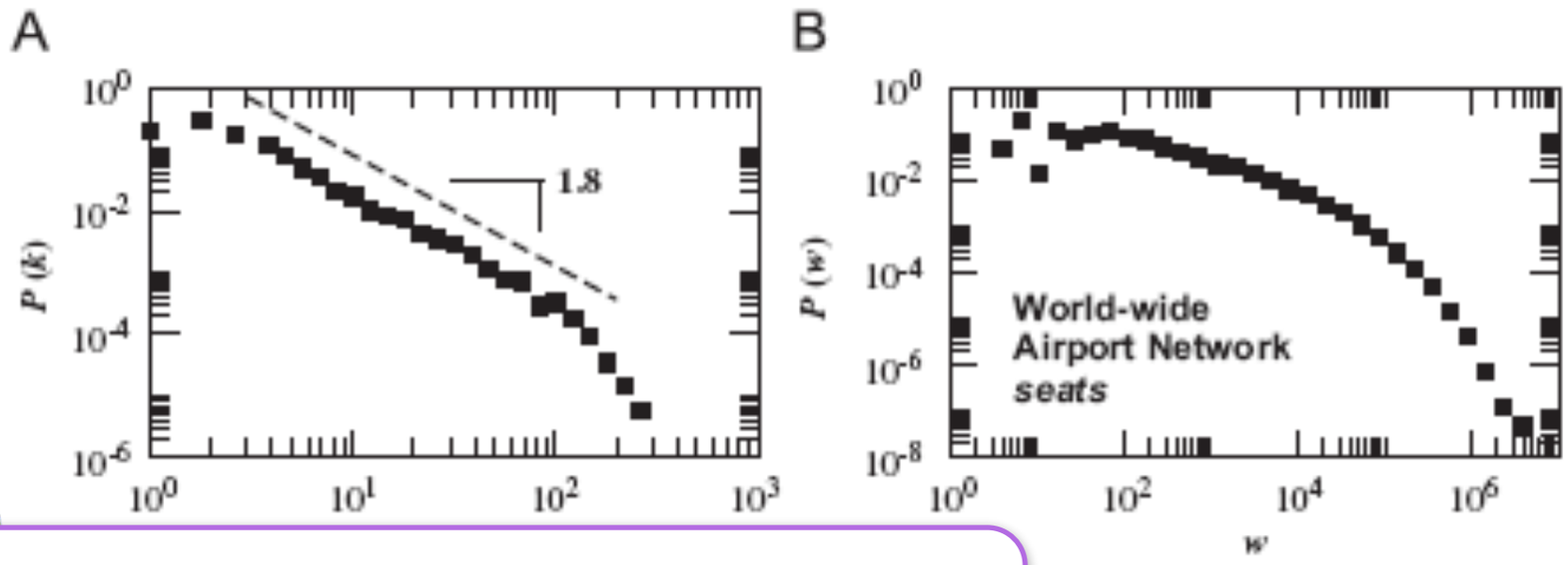
Databases | Summary Tables | Glossary | Filter Subject: All Subjects | Go

<<Prev Rows 1 to 15 of 26 Next>>

Database Name	Description	
Air Carrier Financial Reports (Form 41 Financial Data)	Form 41 Financial Schedule consists of financial information on large U.S. certified air carriers--includes balance sheet, cash flow, employment, income statement, fuel cost and consumption, aircraft operating expenses, and operating expenses. **Note:** Numbers presented on B1, B11 Balance Sheet and P11, P12 Statement of Operations now follow the format of common public financial documents. This format reverses signs from the accounting format in which numbers appeared prior to 10/18/2006 (Examples).	Profile
Air Carrier Statistics (Form 41 Traffic)- U.S. Carriers	Monthly data reported by certificated U.S. air carriers on passengers, freight and mail transported. Also includes aircraft type, service class, available capacity and seats, and aircraft hours ramp-to-ramp and airborne.	Profile
Air Carrier Statistics (Form 41 Traffic)- All Carriers	Monthly data reported by certificated U.S. and foreign air carriers on passengers, freight and mail transported. Also includes aircraft type, service class, available capacity and seats, and aircraft hours ramp-to-ramp and airborne.	Profile
Air Carrier Summary Data (Form 41 and 298C Summary Data)	Summary data of the non-stop segment and on-flight market data reported by air carriers on Form 41 and Form 298C	Profile
Airline On-Time Performance Data	Monthly data reported by US certified air carriers that account for at least one percent of domestic scheduled passenger revenues--includes scheduled and actual arrival and departure times for flights.	Profile
Airline Origin and Destination Survey (DB1B)	Origin and Destination Survey (DB1B) is a 10% sample of airline tickets from reporting carriers. Data includes origin, destination and other itinerary details of passengers transported.	Profile
American Travel Survey (ATS) 1995	National data on the nature and characteristics of long-distance personal travel, from a household survey conducted by BTS about every five years.	Profile

Long Range Mobility

PNAS 106, 21484 (2009)



What about Short Range Mobility?

Complete IATA and OAG databases:

3362 airports worldwide
220 countries
> 20,000 connections
w_{ij} #passengers on connection i-j
>99% total traffic

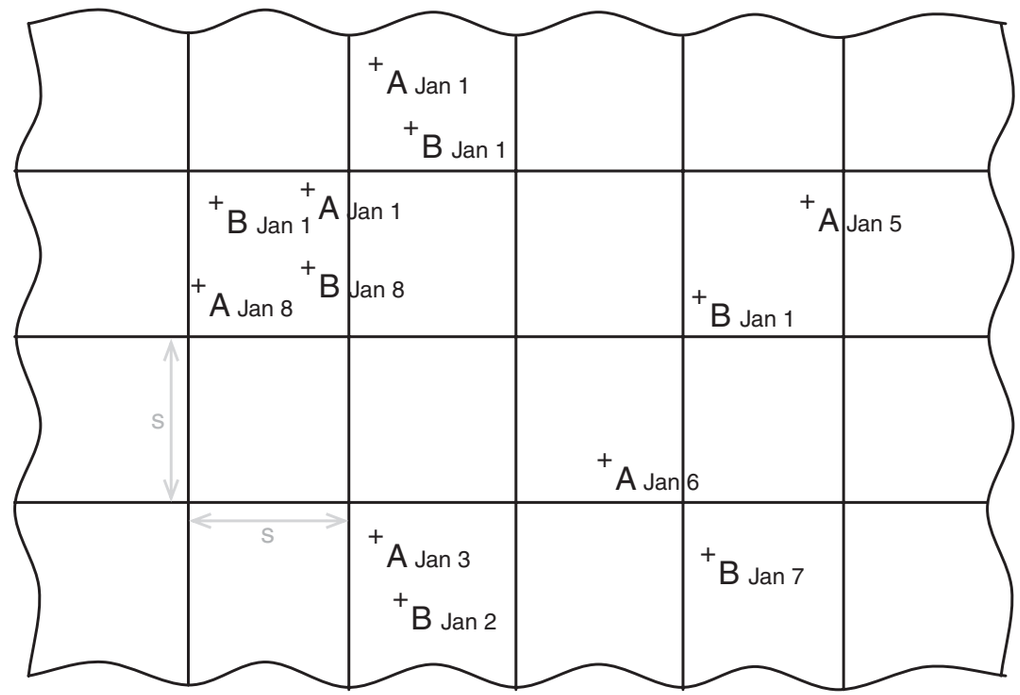
Commuting

Co-occurences and Social Ties

PNAS 107, 22436 (2010)

- Geotagged Flickr Photos

- Divide the world into a grid



Count number of cells on which two individuals were within a given interval

Co-occurences and Social Ties

PNAS 107, 22436 (2010)