

My Life with Pseudogenes

Bonus Feature*

Cristina Sisu

GPMTG

17th April 2014

A life with pseudogenes ...

is amazing for people dealing with them

.... but BORING for those not involved!

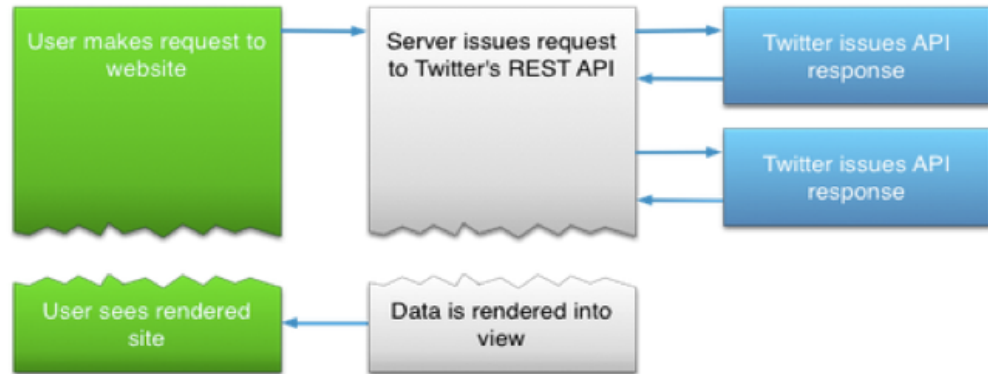
Mark Gerstein's Life on Twitter

Guest Speaker: Dr Ed Cannon



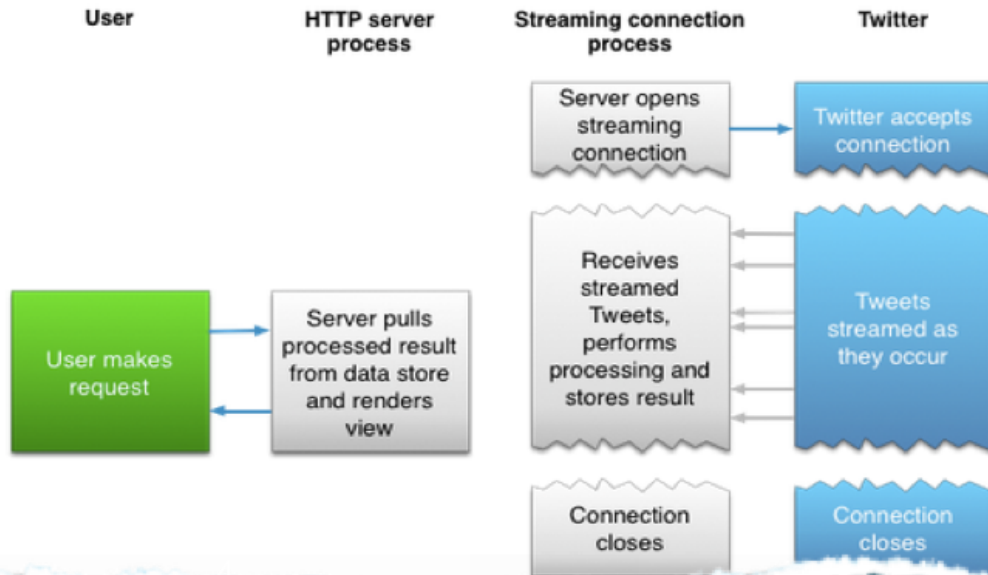
Twitter APIs

- Search API



- Streaming A

An app which connects to the Streaming APIs will not be able to establish a connection in response to a user request as shown in the above example. Instead, the code for maintaining the Streaming connection is typically run in a process separate from the process which handles HTTP requests:



- REST API



Twitter REST API

- Timelines
- Tweets
- Search
- Streaming
- Direct messages
- Friends followers
- Users
- Suggest users
- Favourites
- Lists
- Saved searches
- Places and Geo
- Trends
- Spam reporting
- OAuth
- Help



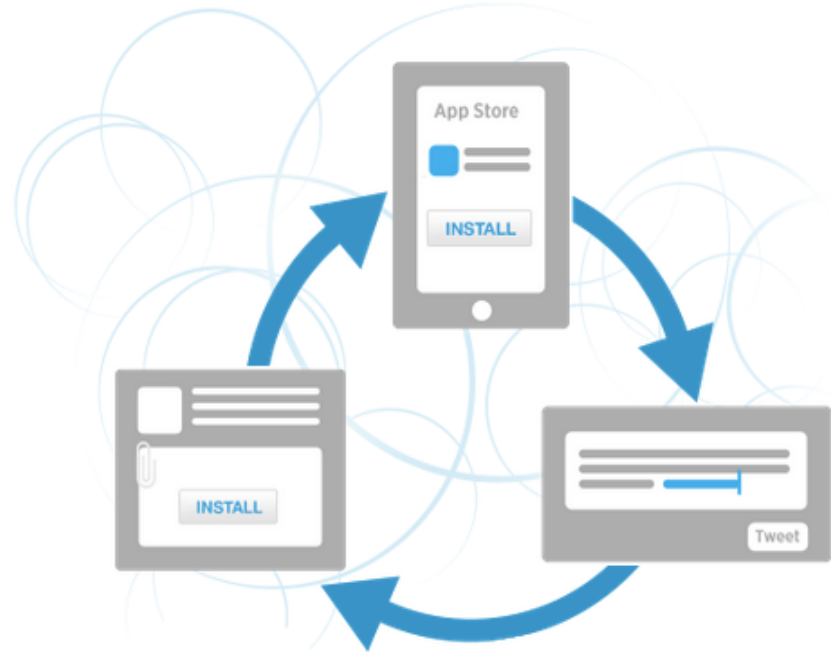


More downloads for your app

with Twitter Cards

Twitter Cards offer a fast and easy way to grow your user base for mobile apps. Simply add some new markup to your pages: when users tweet links to your domain, Cards will let other users viewing those Tweets to download and launch your app across a number of mobile platforms.

[Learn More](#)



Twitter Cards



Embedded Timelines



Embedded Tweets



Tweet Button



Follow Button



https://apps.twitter.com

Application Management

Status

Your application has been successfully deleted.

Twitter Apps

Create New App



MGGroupMeeting

Testing twitter api



Create an application

Application details

Name *

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL

Where should we return after successfully authenticating? [OAuth 1.0a](#) applications should explicitly specify their `oauth_callback` URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.



MGGroupMeeting

Details

Settings

API Keys

Permissions

Application settings

Keep the "API secret" a secret. This key should never be human-readable in your application.

API key

[REDACTED]

API secret

[REDACTED]

Access level

Read-only ([modify app permissions](#))

Owner

csds2eoc21

Owner ID

321605507



Your access token

You haven't authorized this application for your own account yet.

*By creating your access token here, you will have everything you need to make API calls right away.
your application's current permission level.*

Token actions

Create my access token



Accessing the Twitter REST API

Twitter Libraries

Getting Started

View [What links here](#)

Updated on Tue, 2014-01-21 23:42 API version 1.1

These libraries, while not necessarily tested by Twitter, should support Twitter API v1.1.

Want your library to be included in this index or need to update the details we have? [Submit your library for inclusion!](#)

Libraries built and maintained by Twitter

Java

- [hbc](#) — A Java HTTP client for consuming Twitter's Streaming API

Libraries built for the Twitter Platform

Multi-platform

- [Temboo](#) — by @temboo — Framework for working with Twitter via many platforms including iOS, Android, Java, PHP, Python, Ruby, and Node.js

Tags

- [OAuth](#) (179)
- [Libraries](#) (10)
- [Community](#) (34)
- [Javascript](#) (48)
- [ASP](#) (2)
- [Java](#) (24)
- [PHP](#) (44)
- [Ruby](#) (6)
- [.net](#) (5)
- [C#](#) (6)



OAuth

- OAuth 1.0 Protocol published in April 2010

OAuth

From Wikipedia, the free encyclopedia

OAuth is an [open standard](#) for [authorization](#). OAuth provides client applications a 'secure delegated access' to server resources on behalf of a resource owner. It specifies a process for resource owners to authorize third-party access to their server resources without sharing their credentials.



Request token - make signed GET request to the request token URL. (sign request using consumer key and secret)

You get back parameterized authorization url with oauth_token parsed into it e.g. <https://api.twitter.com/oauth/authorize?>

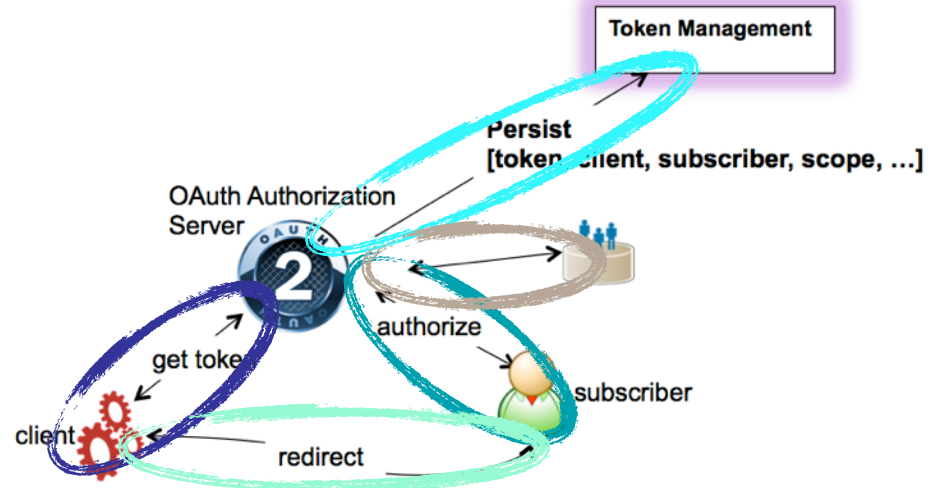
Go to url get users consent to allow app access to your account.

Redirects to app call back url with oauth_token and oauth_verifier.

The oauth_verifier is the access token for the user.

Paste into the code the oauth_verifier

OAuth Handshake



OAuth

Client stores login & password?

✗

Login & Password per request?

✗

Client acquire new password upon change?

✗

Restrict other apps from using your account?

✓

Replayed requests preventable?

✓



Overview: Version 1.1 of the Twitter API

View

[What links here](#)

Updated on Tue, 2013-05-07 07:29

API version 1.1

Version 1.1 of the Twitter API is the first major update of the API since its launch. We're excited about what it means for developers and we've captured all the meaningful changes here so you don't miss a thing.

1. [Improved rate limits](#)
2. [JSON support only](#)
3. [Authentication required on all endpoints](#)
4. [Updated Developer Rules of the Road](#)
5. [Developer Display Requirements](#)
6. [New Twitter client policies](#)
7. [Default entities and retweets](#)
8. [Deprecation of v1.0 of the API](#)
9. [dev.twitter.com Updates](#)
10. [Streaming API changes](#)
11. [Feedback and next steps](#)



... let the OTHERS do the work for you

- <http://www.tweetstats.com/>
- <http://twittercounter.com/>
- <http://twitaholic.com/>
- <http://foller.me/>
- <http://tweetreach.com/>
- <https://sumall.com/20ft>
- <http://www.tweetails.com/>
- and many many more!



It all started on a cold December morning, well afternoon, or maybe it was evening ...



TWEETS **2,158** PHOTOS/VIDEOS **104** FOLLOWING **346** FOLLOWERS **1,261** FAVORITES **2** More ▾



Following

Tweets Tweets and replies

Mark Gerstein

@markgerstein

#Bioinformatics Professor, interested in #DataScience, with #Photography, #Cycling & #History as hobbies

📍 Yale

🌐 Gerstein.info

🕒 Joined December 2010

Tweet to Mark Gerstein



Mark Gerstein @markgerstein · 8h ago

Developmental dynamics... of random monoallelic gene expression. #RNAseq on 100s of instances without causative SNPs cell.com/developmental-...

↩️ ↻️ 1 ⭐ 1 ⋮



Mark Gerstein @markgerstein · Apr 8

Finding lost treasures in #exome sequencing data. Mining off-target, often noncoding, reads from 1000G, TCGA, ESP, &c ncbi.nlm.nih.gov/pubmed/23972387

Who to follow · Refresh · View all



Brocade @BRCDcomm
Followed by BigData Blogs a...

Follow Promoted



Ars Technica @arstechnica

Follow



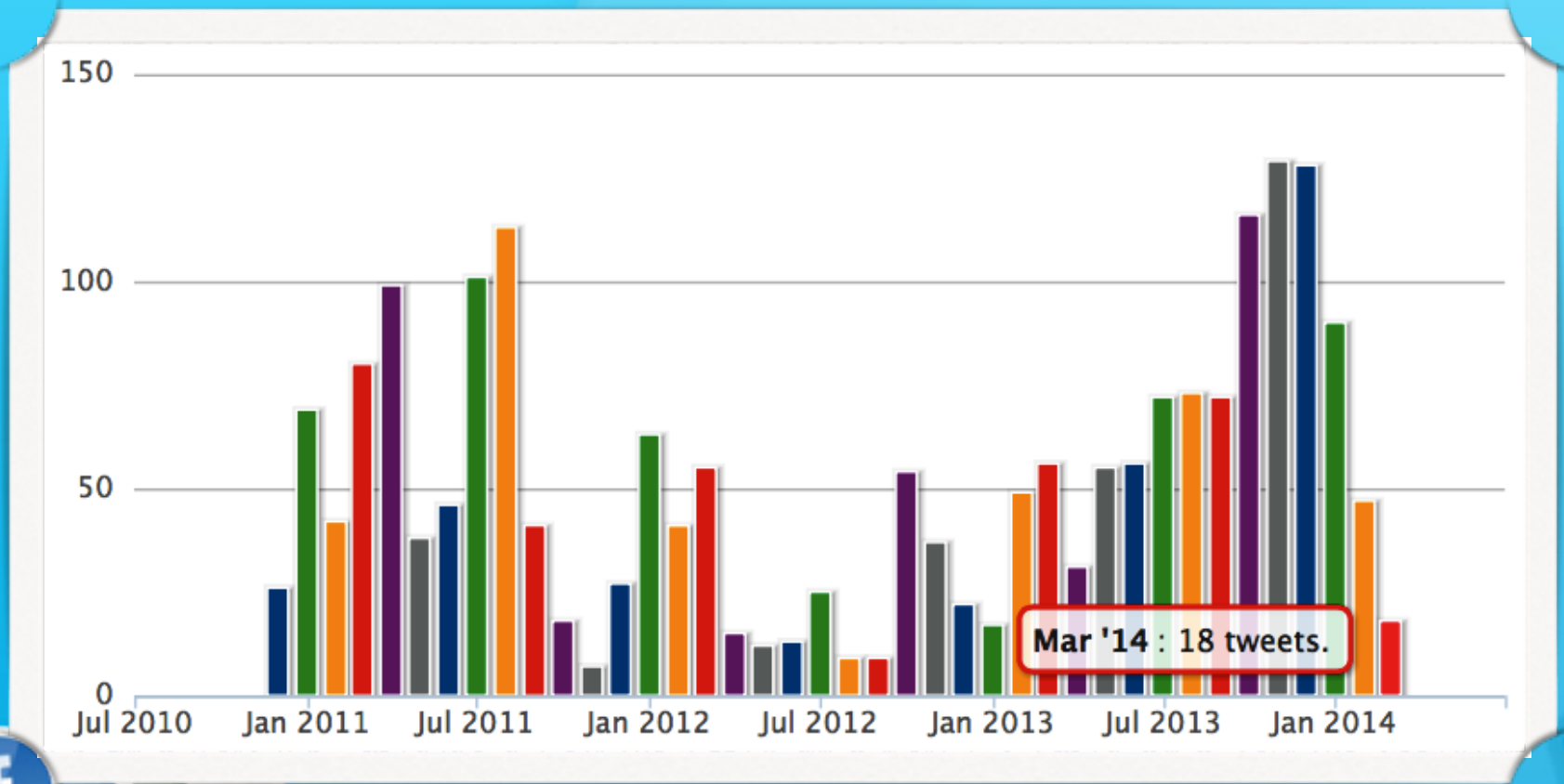
WIRED @WIRED

Follow

Popular accounts · Find friends

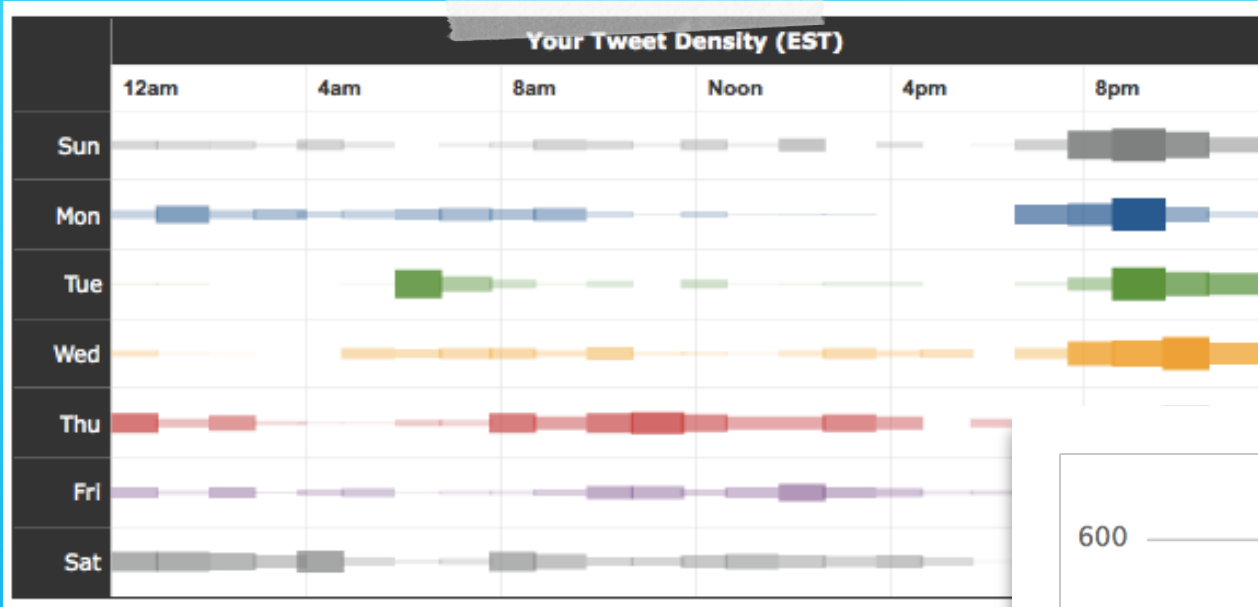


... and since then, @markgerstein has seen highs and lows

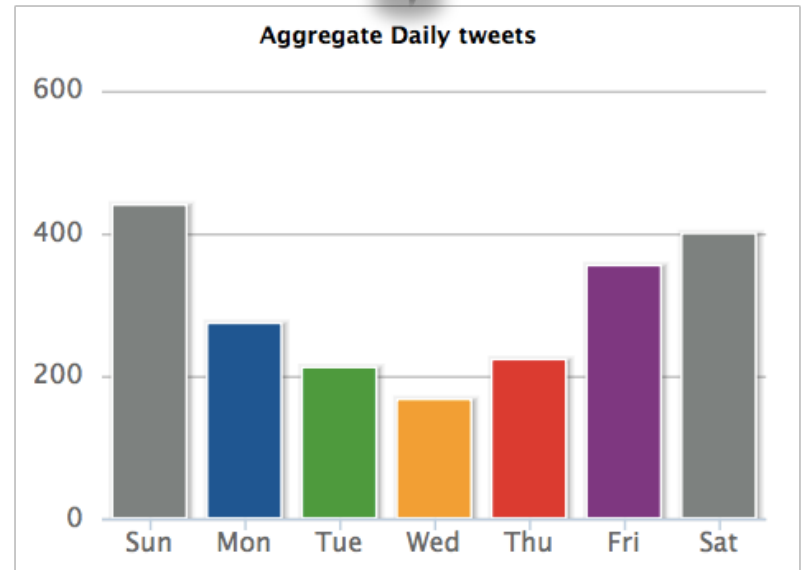


And the iDays are ...

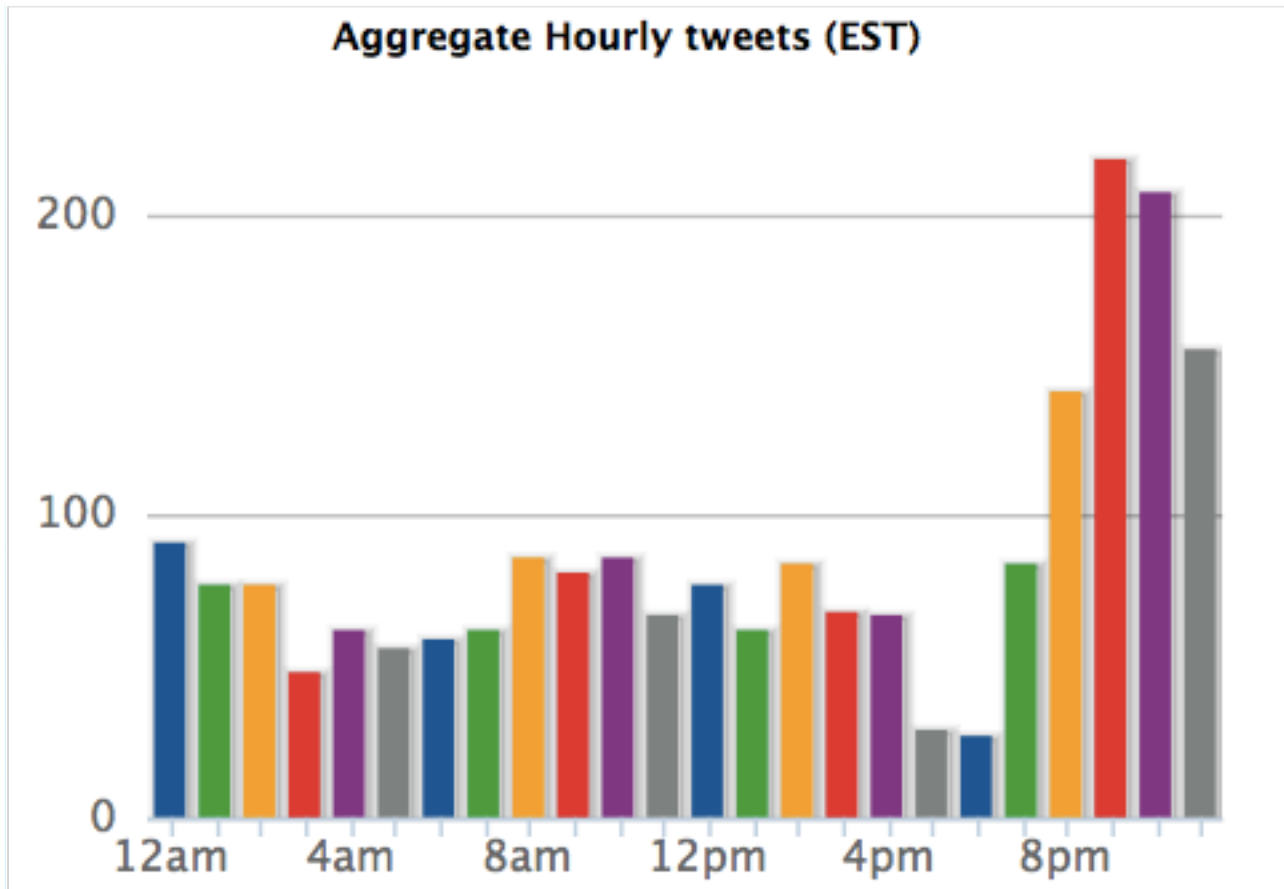
Your Tweet Density (EST)

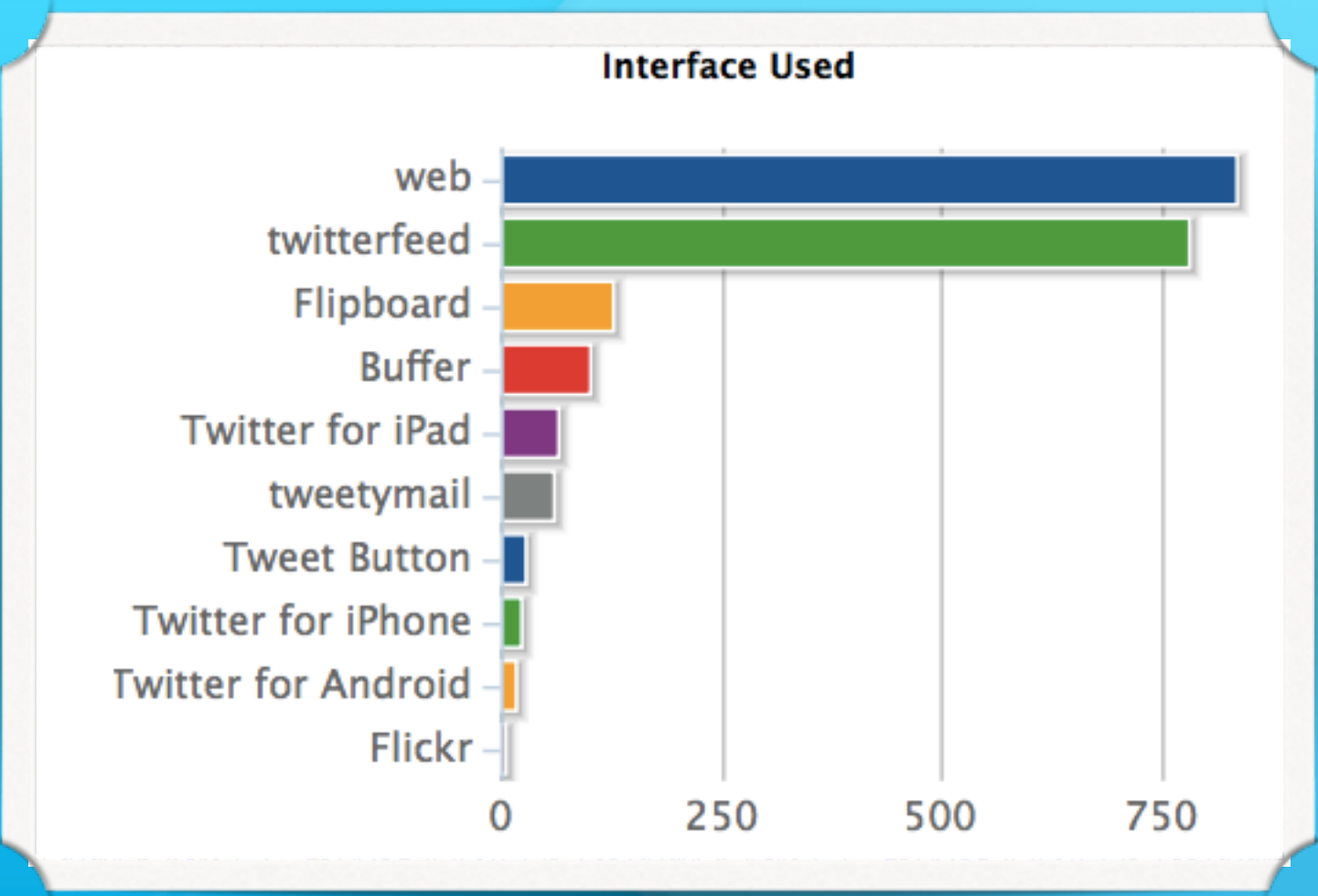


Aggregate Daily tweets



Aggregate Hourly tweets (EST)



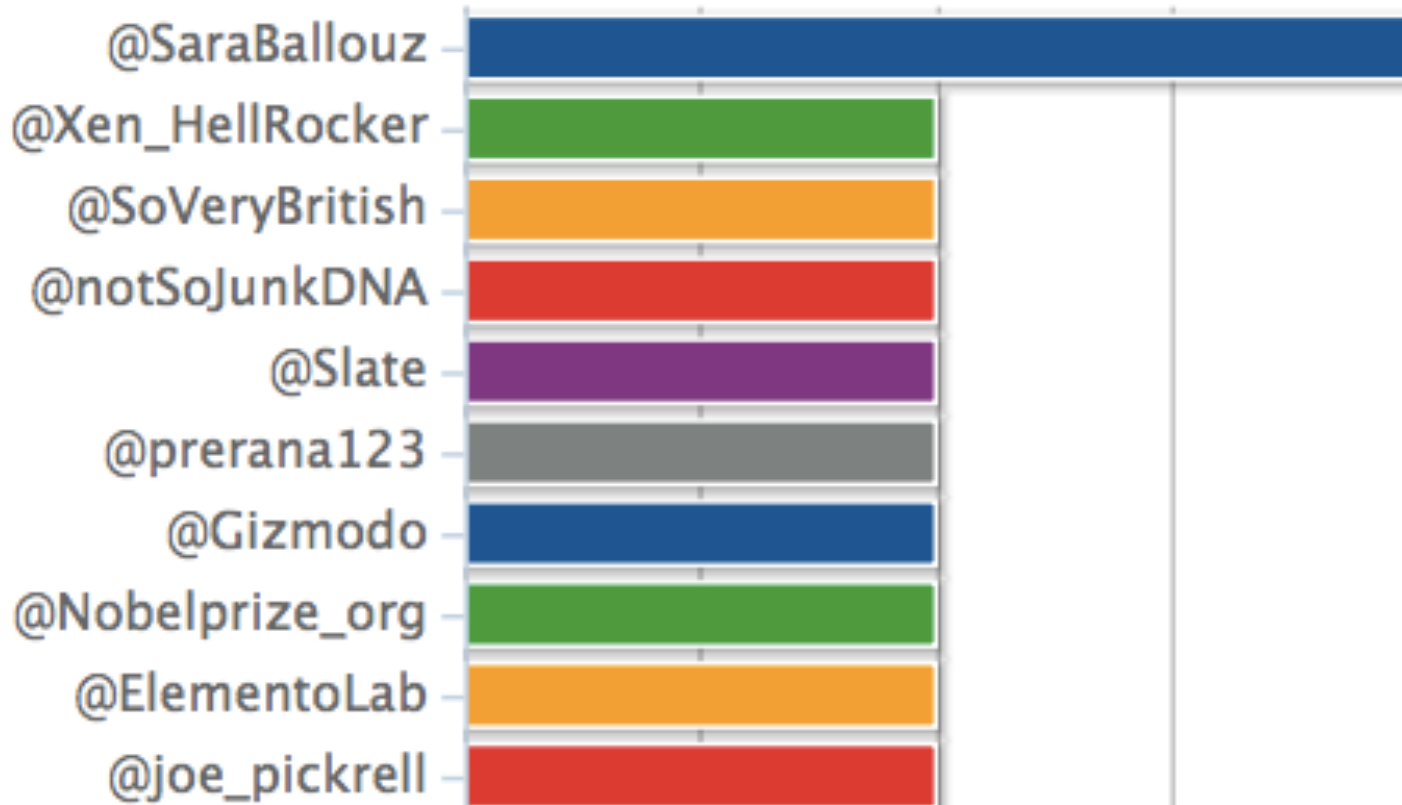


e.g. November 2013 - Tweet Analysis

Y

Replies To (@'s - 6.98% of total tweets)

Who You ReTweet (RT's - 17.83% of total tweets)



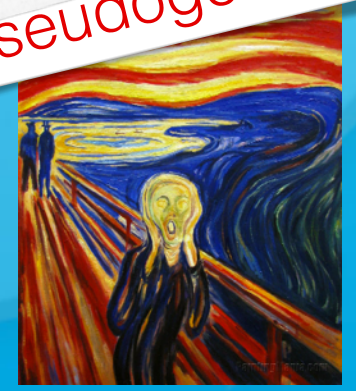
Tweet Word Cloud

center nets history professor
pseudogenes own deal collaborative challenges
guide asthma genetic cupcakes athletic might physical
recent happy terms http little protection wall genomenathan could though
some online protein really shows already shows bayesian even went elementolab chart
ideas flickr waitmossberg carlzimmer need including would issue war down posting sciam transcript
approach photos natural OUT links really links carlzimmer need including would issue war down posting sciam transcript
theorist

#blogpost #cancer #genomics #gi2013 #health #photo #pic #privacy *]] 10 11 2011 [[* amazing american analysis
annotation art article best better big @bioinfotools biology blog book brain cancer @carlzimmer cell complex cover daily
data day @dgmacarthur digital dna does econ **economist** email fall flickr free future gene genes genetic genetics **genome**
genomes **genomics** **good** **google** **great** harvard health history ht human interesting internet @kirkdborne know leaves
letter life lines little long look magazine make material **mbgmbg** @michaelhoffman **ms** **mt** natural **nature** need network
networks **new** **news** @newyorker nice nobel @notsojunkdna ny nyti @nytimes **nytimes** @nytimescience old ones online
open overview @pamefeliciano paper people personal **photo** photos plos @pogue post **posted** power privacy project pt raw
read real really recent research review **rt** says @sciam **science** scientific scientist scientists self sequence sequencing series
sharing shows sky @slate st story study talk tech technology thanks things thoughts time times tricolortimelapse use useful
using variants vs water way web wikipedia work world **wsj** yale years yes york yorker

nytimescom management prize class review patients lots doing gwas
commentary bits with tags markgerstein been chriswiggins column makes iphone video still internet digital
email mentioned stats stanford scientist editor limits mining has
glass design name encyclopedia printer imagine provost bring genomes elements under mims relating alive
predict check deep kirkdborne only software night cancer researchers runningintotheoilingseadsc thoughtful now computer
interpretation mapping differences ewanbirney personalized runningintotheoilingseadsc thoughtful now computer
expression newyorker before personalized runningintotheoilingseadsc thoughtful now computer
figer plays notsojunkdna atulbutte problem

No Pseudogenes!



A bit of tweet-detail

- Space efficiency: 79.9% — tweets have on average 28 spare characters left over
(tweet max character count = 140)
- Average word length: 4.97 characters — one character longer than M.A.R.K :)
- Vocabulary diversity score: 47.63/100 — could improve on that one
- Real dictionary words: 88.98%
- Activity: 1.78 tweets/day (average person post 0.3 tweets/day!)







tweetails.com resolution on Mark:

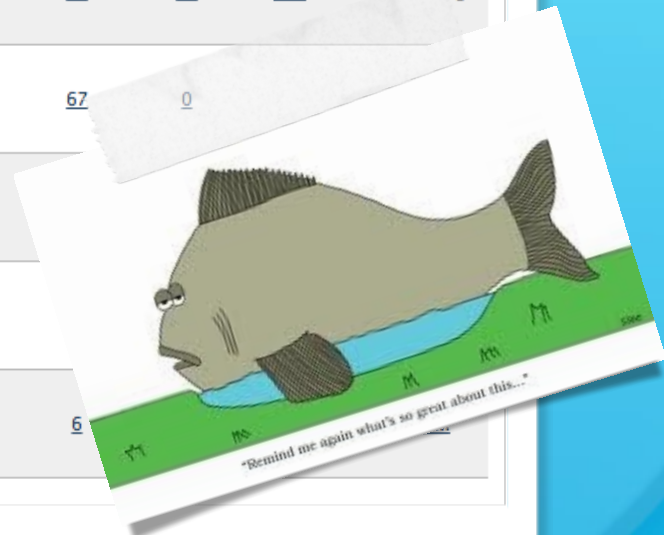
Our intelligence rating for Markgerstein is: -

Superhuman - *Markgerstein is so smart they are clearly a member of a super-intelligent alien race posing as a human. Don't worry, we've alerted the authorities!*






So where does Mark stand on the Twitter scale?

#	Name (Screen Name)	Location	URL	Followers	Following	Updates	Joined
1.	 Mark Gerstein @markgerstein	Yale	http://t.co/Fb9oq1ZqXT	<u>1,133</u>	<u>336</u>	<u>2,077</u>	Details...
2.	 John @CanWeBowlPlease	Yale		<u>76</u>	<u>55</u>	<u>867</u>	78 months ago
3.	 John @CanWeBowlPlease	Yale		<u>67</u>	<u>0</u>		
4.	 n @nzle	yale	http://offtherails.wordpress.com				
5.	 WYBCX @wybcx	Yale	http://wybcx.com				
6.	 Yan Zhang @supercider100	Yale	https://t.co/Vlad2wd4IR	<u>6</u>			



The outreach


Tweets most retweeted

 **Mark Gerstein** @markgerstein August 15, 2013, 1:59 pm via web  26  3
Gersteinlab.org/Jobs : #Bioinformatics postdoc positions @Yale. Research on #personalgenomics, #networks, nextgen sequencing, #datascience




 **Mark Gerstein** @ma
Bring on the box plots:
to create them blogs.n

 **Mark Gerstein** @ma
Mapping rare & comm
yang of #RVAS v #GW

 **Mark Gerstein** @ma
Posting my #gi2013 t
#Cancer #Genomics

 **Mark Gerstein** @ma
What is a support vect
#teaching! nature.com




★ Tweets most favorited

 **Mark Gerstein** @markgerstein February 5, 2014, 3:54 am via web  12  22
What is a support vector machine overview w/o equations, just pictures. Great for

 **Mark Gerstein** @markgerstein
... via Buffer  24  11
... #plots, solicits & then publishes an R #tool

 **Mark Gerstein** @markgerstein
... am via web  16  11
... an diseases: great primer, describing yin &

 **Mark Gerstein** @markgerstein
... 2014, 4:01 am via web  10  10
MT @... what changed the world http://t.co/9zlfJPxFFJ Great list but
would've included S&K inv w/ dS>0 in eq#12

 **Mark Gerstein** @markgerstein March 20, 2014, 3:20 pm via Buffer  6  7
Posting #talk @nygenome: Human #Genome Analysis: Building the Reg. #Network & Using it to
Interpret #Cancer lectures.gersteinlab.org/summary/Genome...

3,694

accounts reached



Who are Mark's
followers?

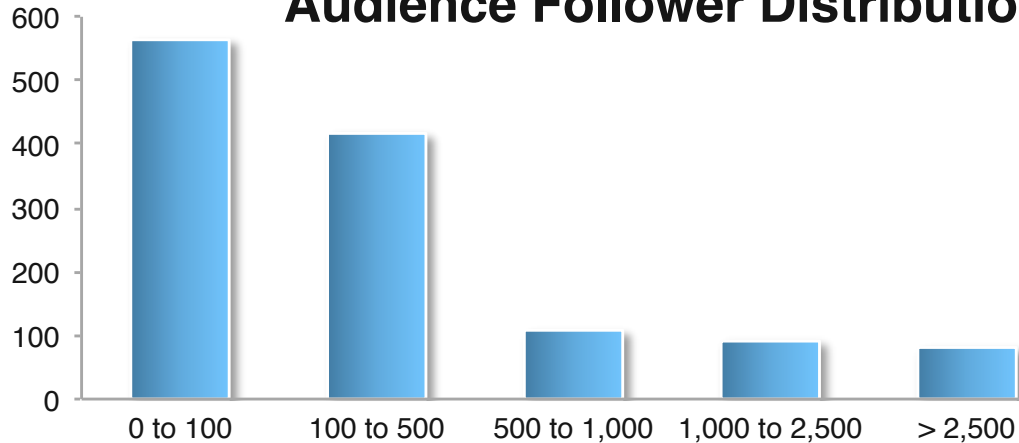


Mark vs Followers

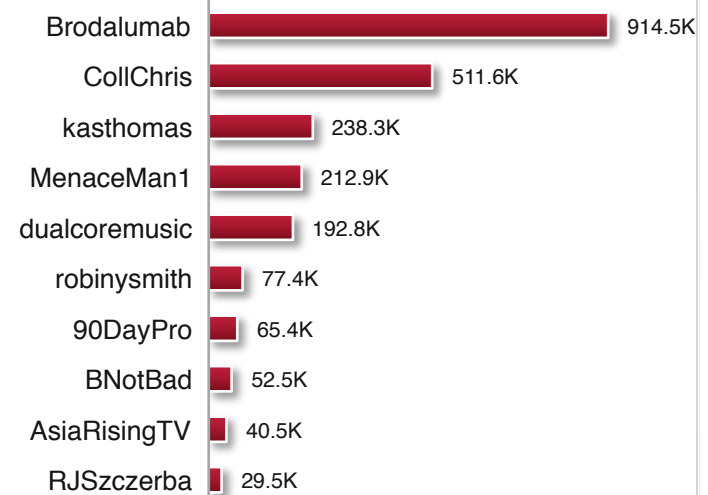
- 1261 people/accounts are following @markgerstein
- ~2508 people/accounts are following Mark's followers

Followers win!

Audience Follower Distribution



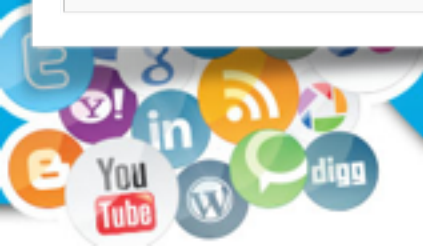
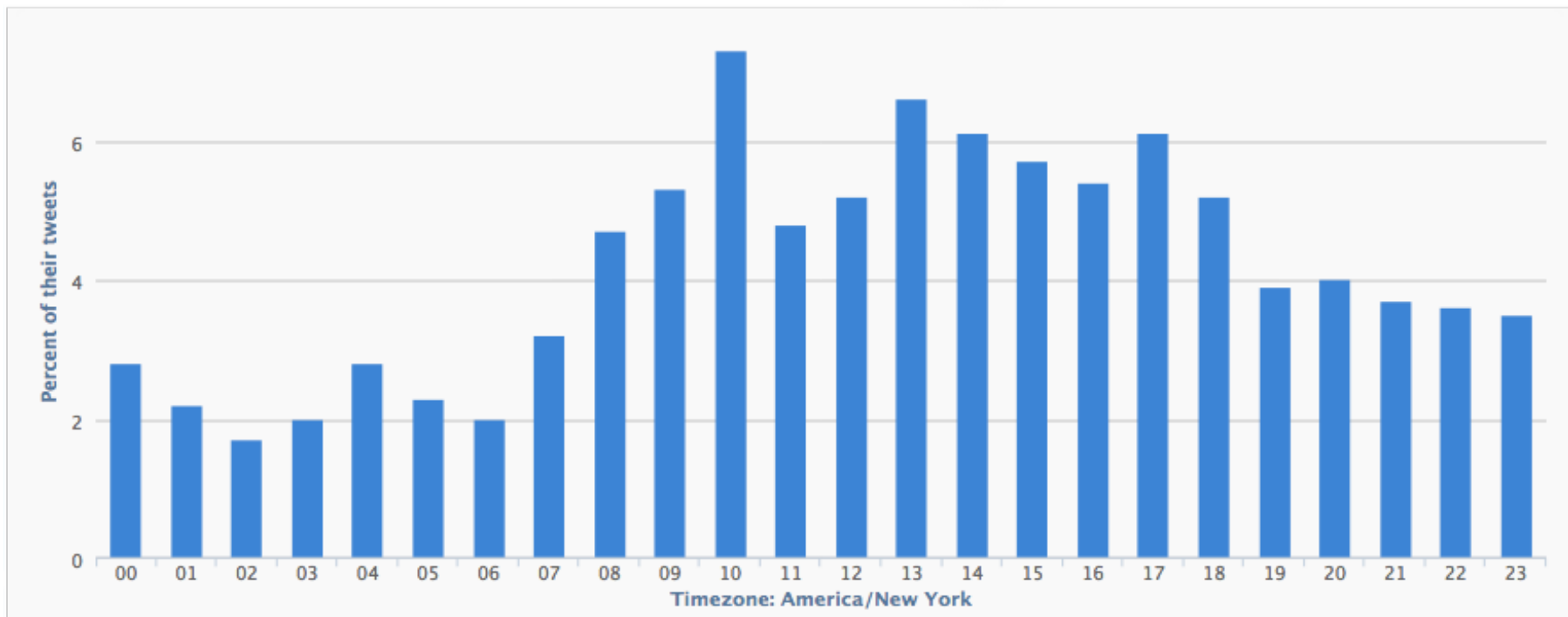
Top Followers by # Followers



Mark tweets at night - his followers tweet ...

- Mark tweeted 2158 times
- Mark's followers tweeted 1635 times

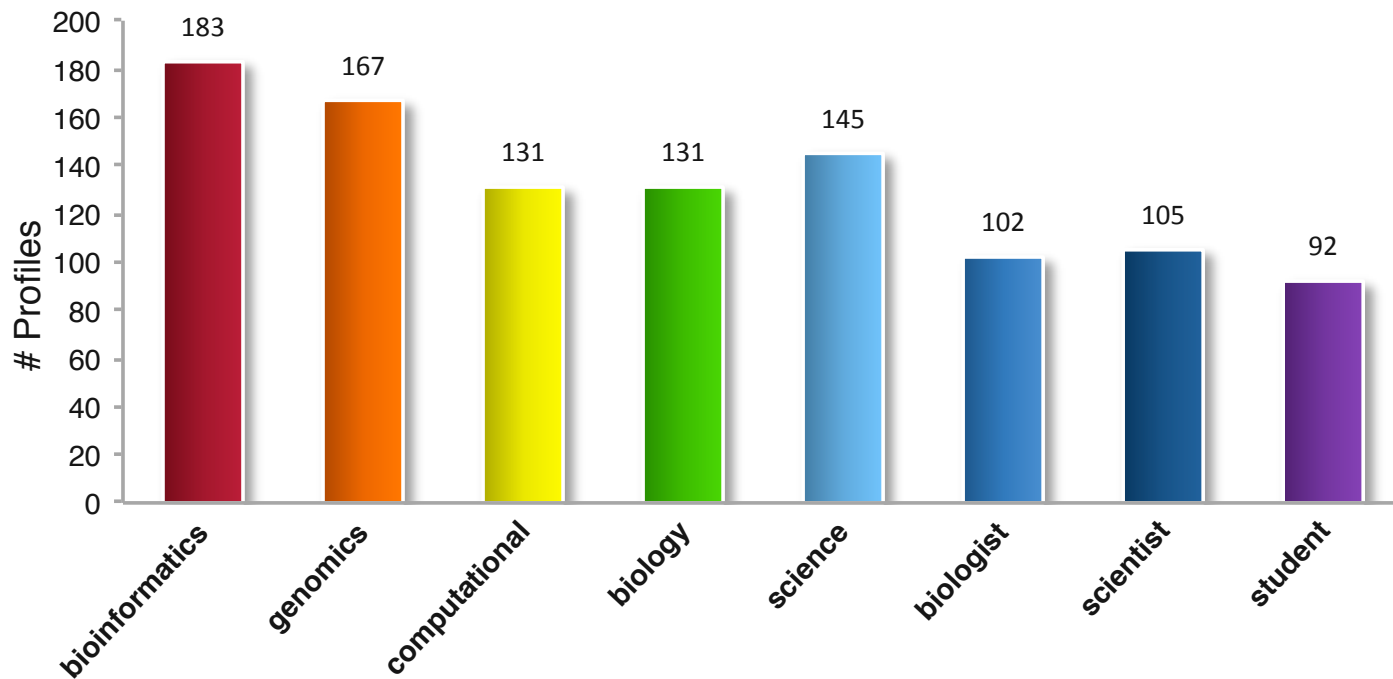
Mark wins!



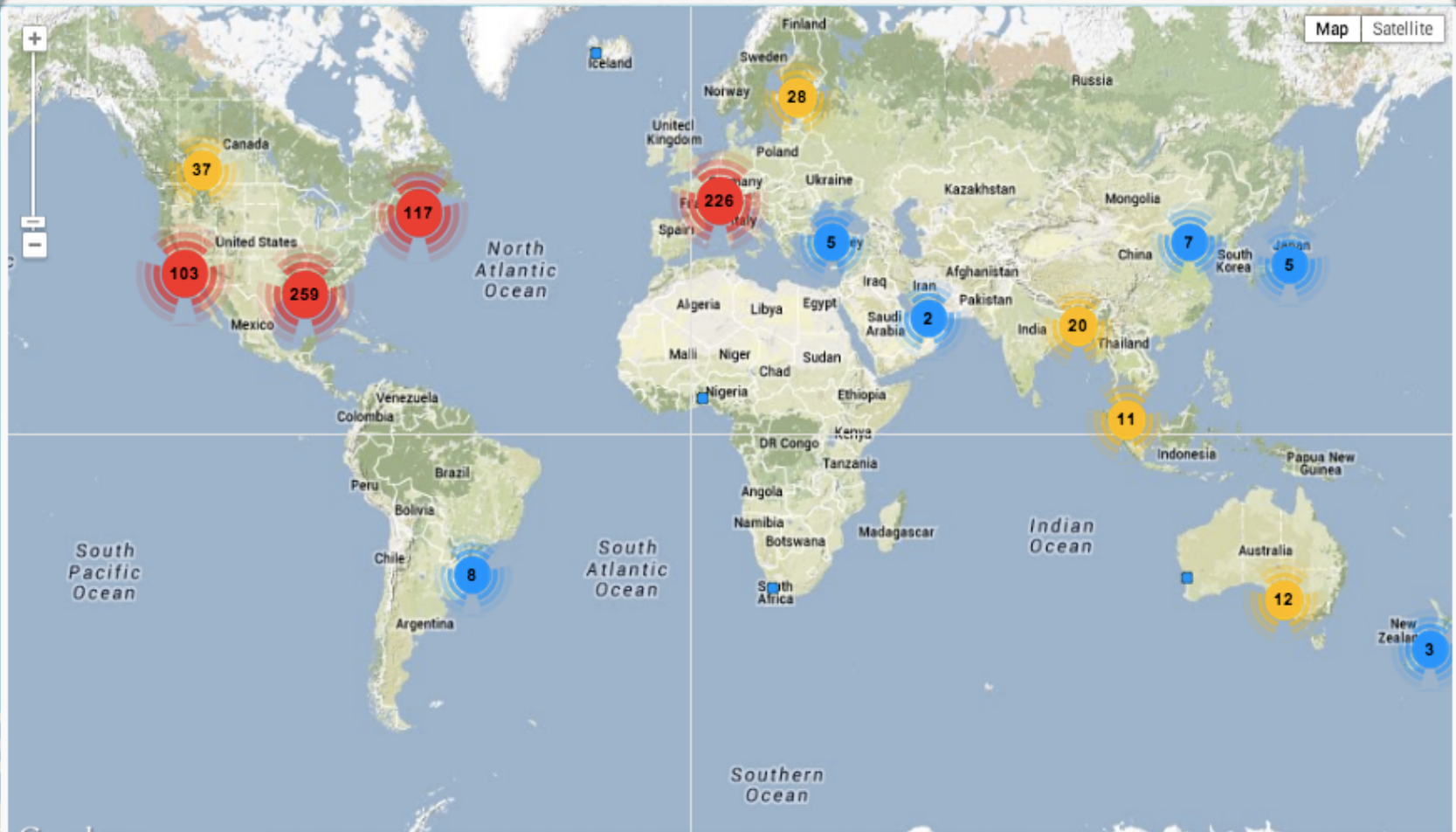
... during the day!

So who are Mark's followers?

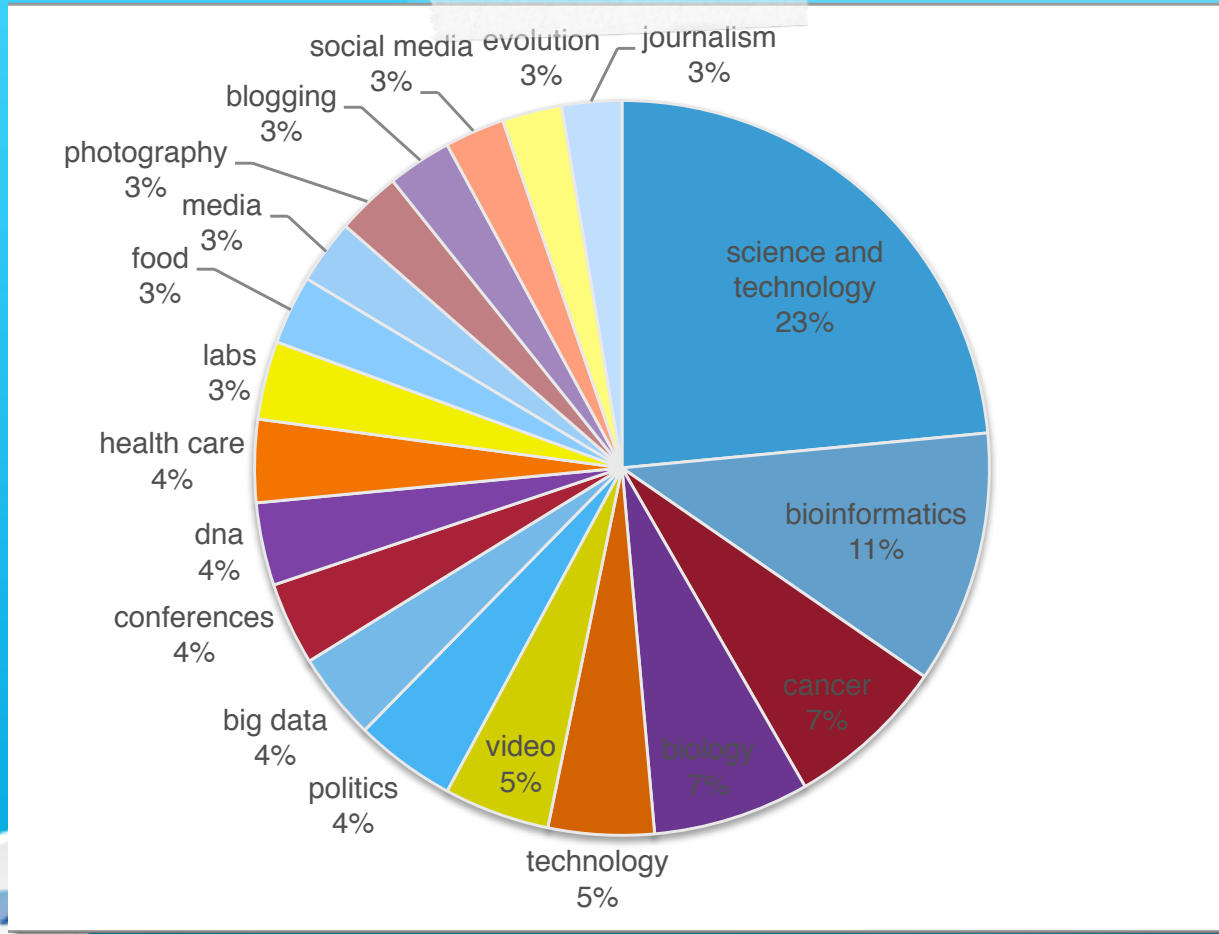
Top keywords in followers descriptions



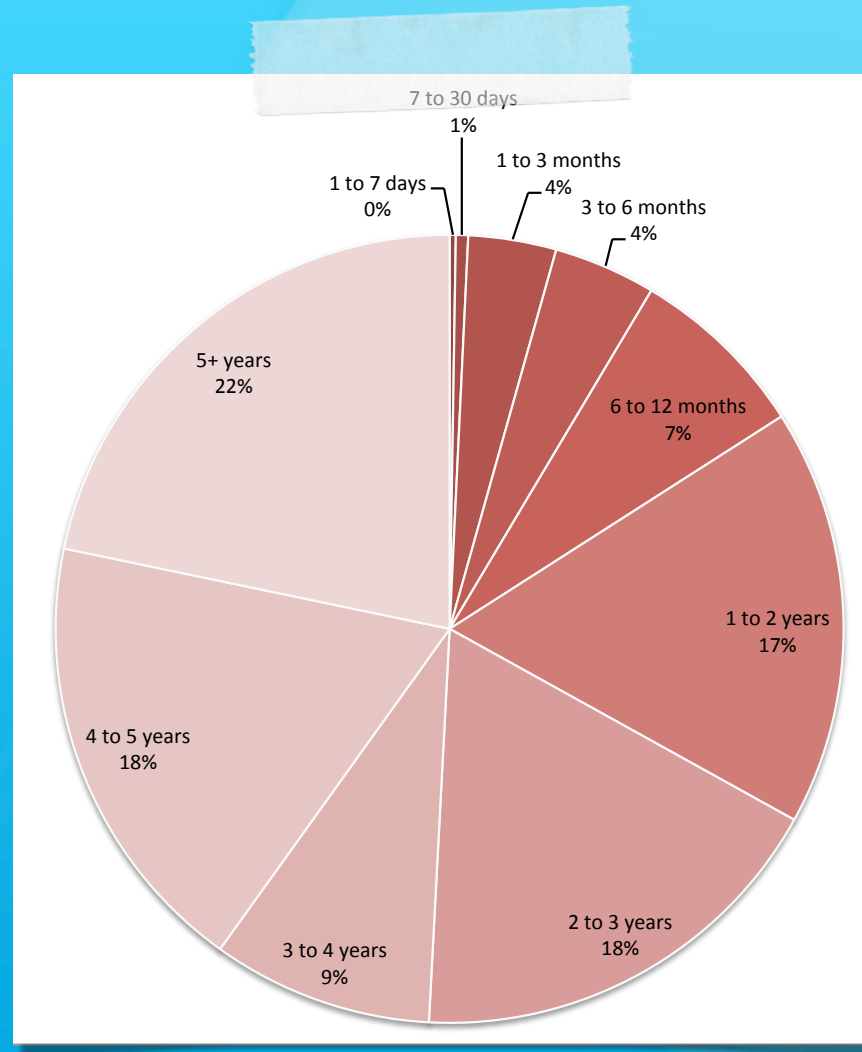
On the map ...



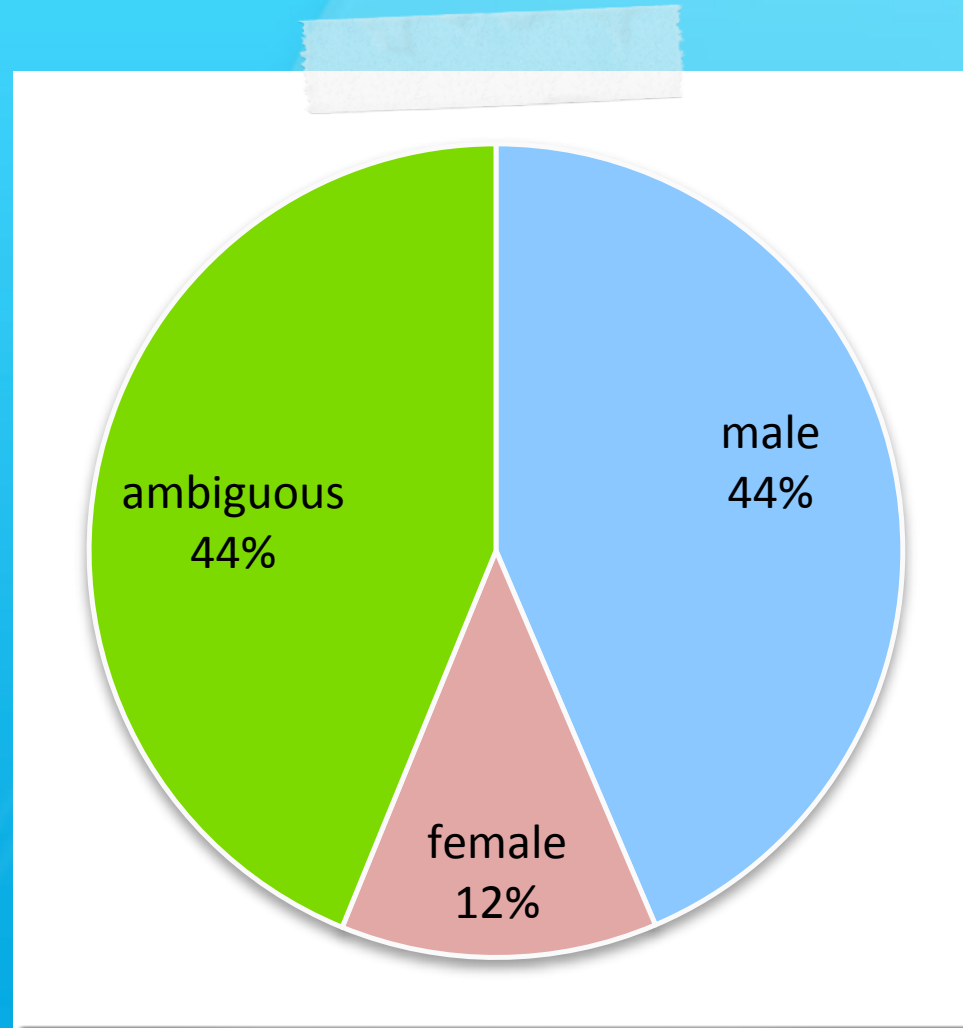
What's the talk all about ?



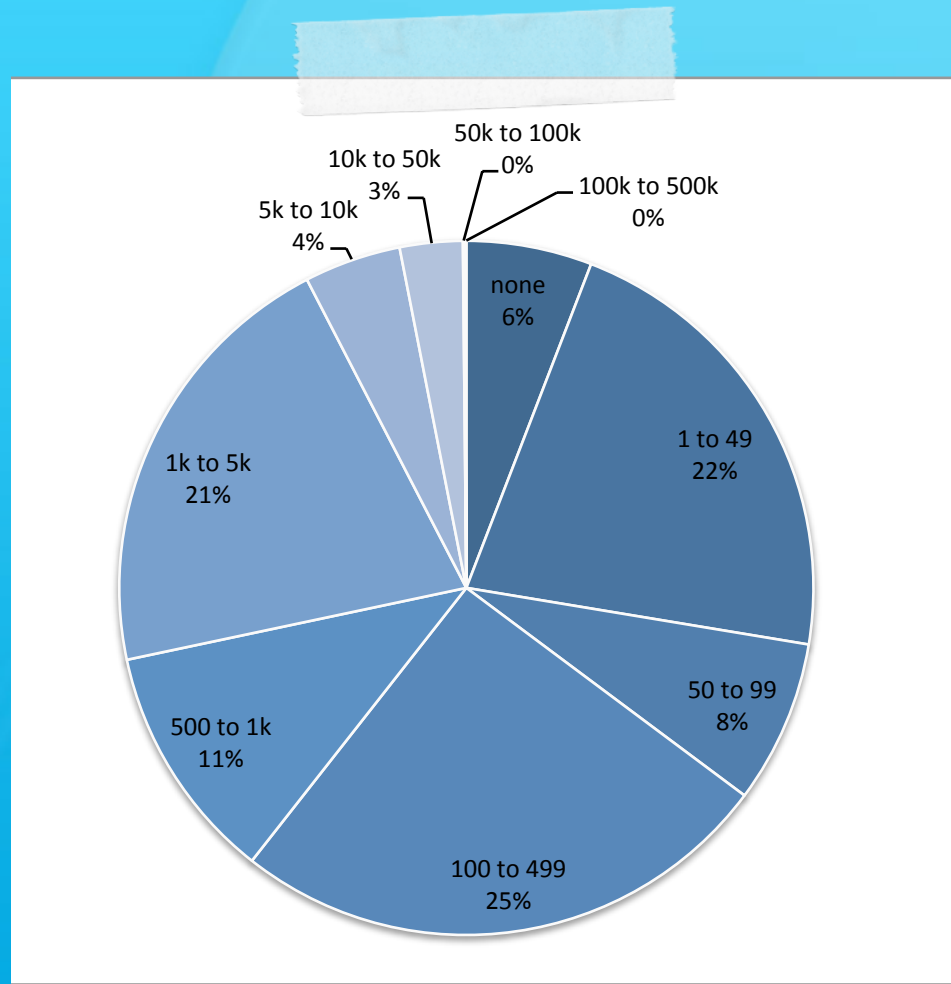
Account ages of Mark's followers



Gender of Mark's followers



Total tweets of Mark's followers













Hall of Fame

↕ Users most retweeted

 @nytimes	16	 @sciam	15
 @NewYorker	14	 @carlzimmer	13
 @dgmacarthur	11	 @Pogue	11
 @Magda_Skipper	9	 @pamefeliciano	8
 @waltmossberg	8	 @KirkDBorne	7

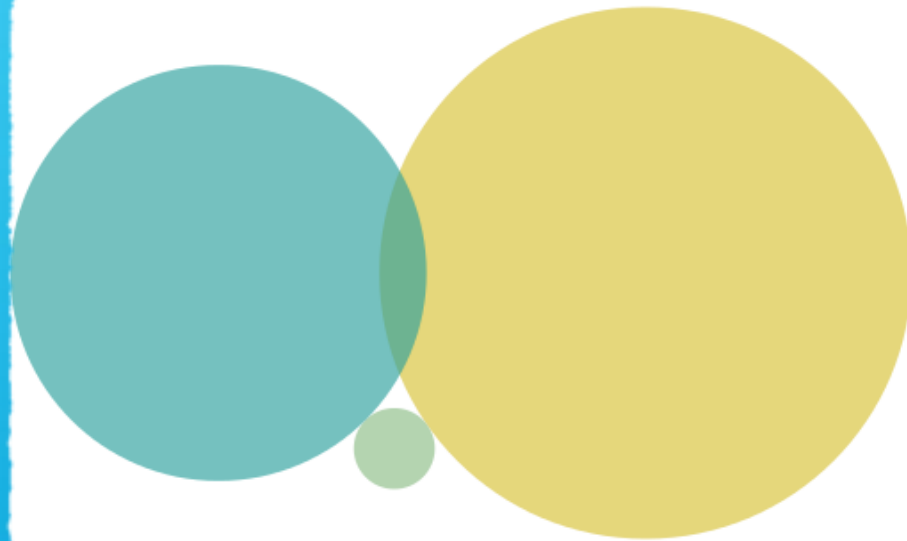
💬 Users most replied to

 @markgerstein	18	 @BioinfoTools	12
 @SaraBallouz	7	 @iddux	7
 @michaelhoffman	6	 @wjrl59	5
 @ewanbirney	5	 @timtriche	5
 @jflier	4	 @KirkDBorne	4



Go compare!

Comparison of users markgerstein & Beyonce & eoc21 follow



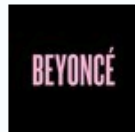
- 345 users followed by markgerstein
- 8 users followed by Beyonce
- 211 users followed by eoc21

followed only by markgerstein »	335	60.5%
followed only by Beyonce »	8	1.4%
followed only by eoc21 »	201	36.3%
followed only by markgerstein & Beyonce	0	0.0%
followed only by markgerstein & eoc21 »	10	1.8%
followed only by Beyonce & eoc21	0	0.0%
followed by all three	0	0.0%
<i>combined total followed</i>	554	



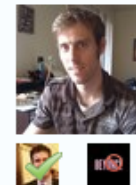
Mark Gerstein

Social Authority: 25
 Followers: 1,264
 Days on: 1,201



Beyoncé Knowles

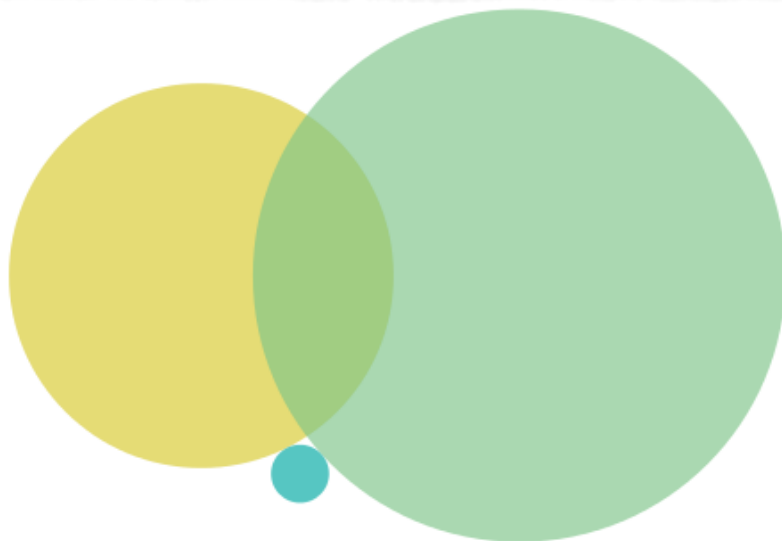
Social Authority: 63
 Followers: 13,331,924
 Days on: 1,823
 Retweets: 0.0%
 @Contacts: 0.0%
 URL tweets: 75.0%



Ed Cannon

Social Authority: 1
 Followers: 50
 Days on: 1,704





- 351 users followed by markgerstein
- 674 users followed by dgmacarthur
- 8 users followed by Beyonce

followed only by markgerstein »	256	27.3%
followed only by dgmacarthur »	579	61.7%
followed only by Beyonce »	8	0.9%
followed only by markgerstein & dgmacarthur »	95	10.1%
followed only by markgerstein & Beyonce	0	0.0%
followed only by dgmacarthur & Beyonce	0	0.0%
followed by all three	0	0.0%
<i>combined total followed</i>	938	

Mark Gerstein

Social Authority: 25

Followers: 1,270

Days on: 1,205

Daniel MacArthur

Social Authority: 60

Followers: 12,701

Days on: 2,016

Retweets: 37.5%

@Contacts: 31.0%

URL tweets: 19.5%

Beyoncé Knowles

Social Authority: 63

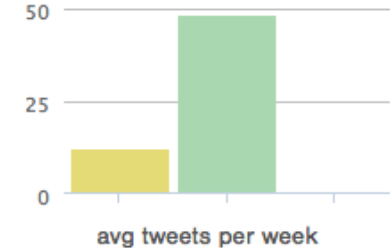
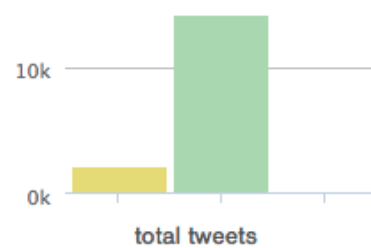
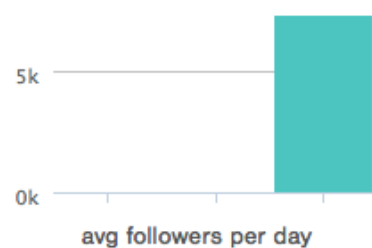
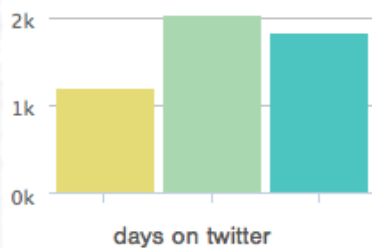
Followers: 13,335,101

Days on: 1,827

Retweets: 0.0%

@Contacts: 0.0%

URL tweets: 75.0%



Conclusions

- There is a lot of public information out there that can be used in various ways ...
- ... better keep with pseudogenes!



Acknowledgements

Ed Cannon

Twitter

All the twitter data analysts out there!

