



**BC COMS 2710:**  
**Computational Text Analysis**

**Lecture 13 – Data Collection**  
**API's - Twitter**



1. Create a Twitter developer account <https://developer.twitter.com/>
2. Go to <https://developer.twitter.com/en/apps> and log in with your Twitter user account.
3. Click “Create an app”
4. Fill out the form, and click “Create”
5. A pop up window will appear for reviewing Developer Terms. Click the “Create” button again.

Instructions from <http://socialmedia-class.org/twittertutorial.html>



- Readings 04:
  - link posted to course site later today
  - due Sunday
  
- HW 02:
  - Due Wednesday night
  
- Tutorial 4.1:
  - Releasing tonight
  - Topic Modeling for yesterday's data
  
- Assignments going forward



- Mid-semester anonymous brief survey
- What have you learned so far and how comfortable do you feel with the material?
- What has been going well in the course so far? What are things you are enjoying about the course?
- What has not been going well in the course so far? What are things you are not enjoying about the course?
- What can we (the course staff) be doing better?



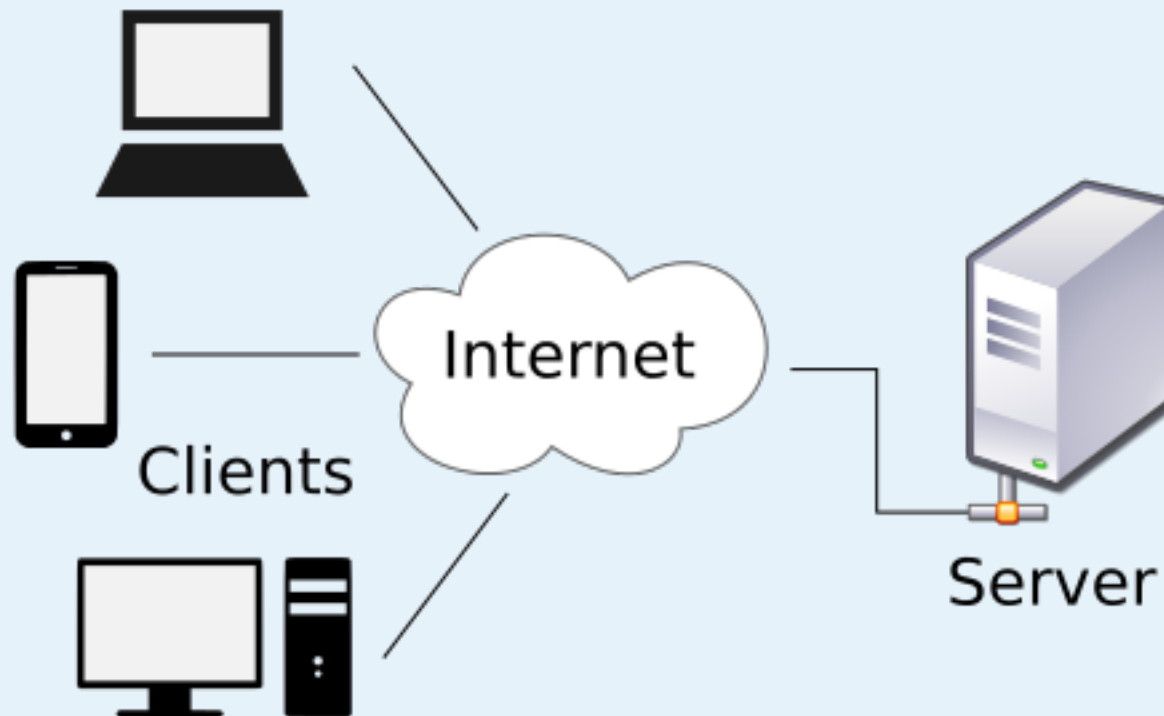
- **Project ideation – Friday May 28<sup>st</sup>**
  - <https://www.overleaf.com/read/yzpgxcgsqdvj>
- roughly 250 word overview of what you are interested in





# Application Programming Interface

# Yesterday: HTTP





# Interacting with servers (via the internet)



## Without API:

An app finds the current weather in London by opening

<http://www.weather.com/> and reading the webpage like a human does, interpreting the content.

Chris Beach



# Interacting with servers (via the internet)



## Without API:

An app finds the current weather in London by opening

<http://www.weather.com> and reading the webpage like a human does, interpreting the content.

## With API:

An app finds the current weather in London by sending a message to the

[weather.com](http://www.weather.com) API (in a structured format like JSON). The [weather.com](http://www.weather.com) API then replies with a structured response.

Chris Beach



Generally: set of protocols that specify how software programs communicate with each other

Programmatically extract and interact with data

Released by companies for data sharing purposes



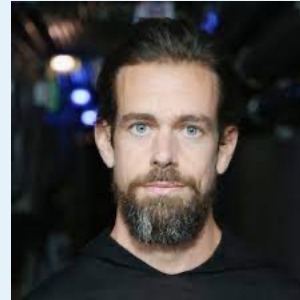


# Twitter

Slides from Alan Ritter



- Jack Dorsey's idea (a NYU undergrad then)
- 1<sup>st</sup> Tweet on March 21, 2006
- 100m tweets/quarter in 2008  
50m tweets/day in 2010,  
400m tweets/day in 2013







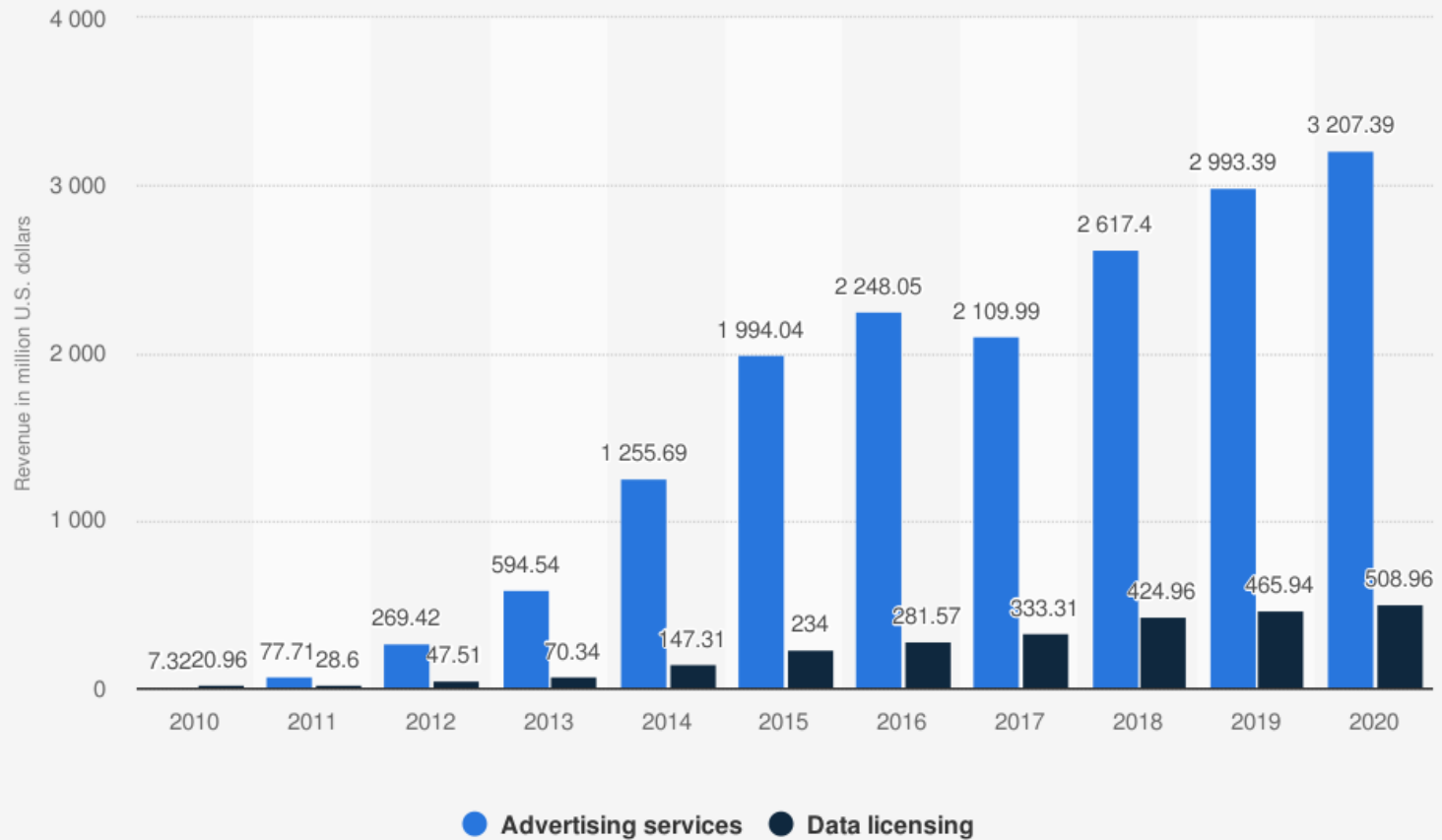
- IPO in 2013 (4<sup>th</sup> quarter)
- Stock overall hasn't done great but recent resurgence
- CEO Dick Costolo resigned July 2015
- Dorsey permanent CEO since October 2015



# Twitter Revenue



Annual revenue of Twitter from 2010 to 2020, by segment (in million U.S. dollars)



Source  
Twitter  
© Statista 2021

Additional Information:  
Worldwide; Twitter; 2010 to 2020






# What's in a Tweet





**President Biden** 

@POTUS

 United States government official



Since we took office, hunger rates have dropped 43% among adults. They're at their lowest rates since the start of the pandemic.

That's the American Rescue Plan at work.

1:30 PM · May 23, 2021 · The White House

**8,789** Retweets   **756** Quote Tweets   **73.2K** Likes






# Tweet Character Limits




**jack**  @jack · Sep 26, 2017 ...

This is a small change, but a big move for us. 140 was an arbitrary choice based on the 160 character SMS limit. Proud of how thoughtful the team has been in solving a real problem people have when trying to tweet. And at the same time maintaining our brevity, speed, and essence!

 **Twitter**  @Twitter · Sep 26, 2017

Can't fit your Tweet into 140 characters? 

We're trying something new with a small group, and increasing the character limit to 280! Excited about the possibilities? Read our blog to find out how it all adds up. 

[cards.twitter.com/cards/gsbby/4ub...](https://cards.twitter.com/cards/gsbby/4ub...)

 6.3K


 33.1K

 24.7K



# ReTweet – re-posting someone else’s tweet



**President Biden** 

@POTUS

 United States government official



Since we took office, hunger rates have dropped 43% among adults. They’re at their lowest rates since the start of the pandemic.

That’s the American Rescue Plan at work.

1:30 PM · May 23, 2021 · The White House

**8,789** Retweets

**756** Quote Tweets

**73.2K** Likes





**Keegan-Michael Key**  @KeeganMKey · May 6, 2020 ...

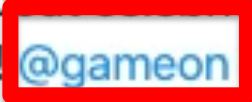

Looking for sports? We have them! How about celebrities and athletes battling it out in epic challenges? We have that too! @gameon Coming May 27th on @cbstv @gronk @venuswilliams @bobbyleelive @iankarmel  
[twitter.com/gameon/status/..](https://twitter.com/gameon/status/..)

- Shortened for display
- Can provide external information for text processing

# Mentions



**Keegan-Michael Key**  @KeeganMKey · May 6, 2020 ...

Looking for sports? We have them! How about celebrities and athletes battling it out in epic challenges? We have that too!  @gameon Coming May 27th on @cbstv @gronk  @venuswilliams @bobbyleelive @iankarmel [twitter.com/gameon/status/...](https://twitter.com/gameon/status/...)



# Replies/Conversations

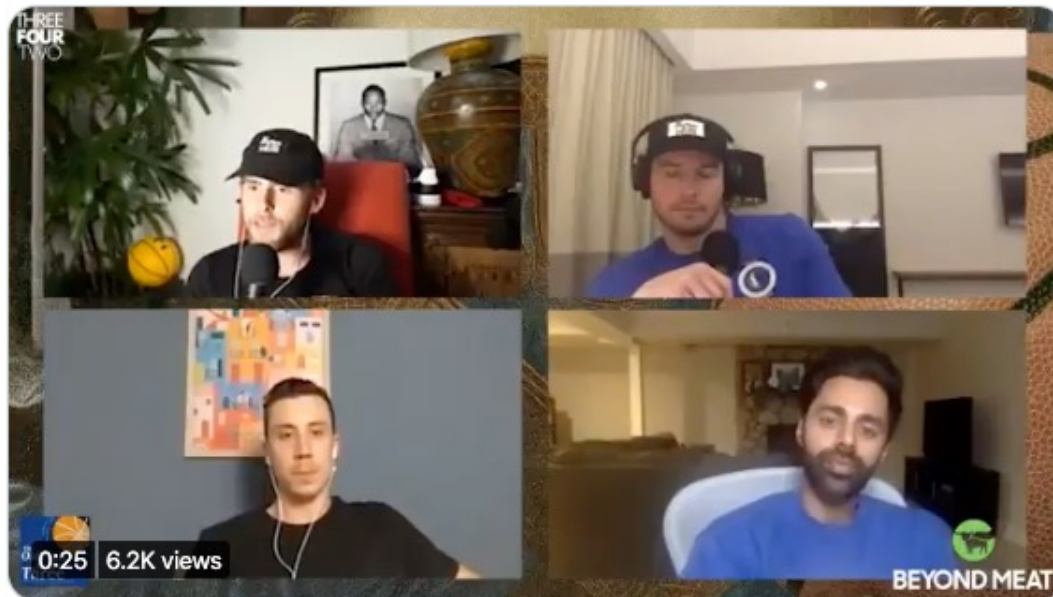


**TheOldMan&TheThree** ✓

@OldManAndThree



On New York basketball. This is a special duo. Episode 55 with [@hasanminhaj](#) and [@D\\_Bo20](#) live now.



10:50 AM · Apr 29, 2021 · Twitter for iPhone

5 Retweets 3 Quote Tweets 71 Likes



# Replies/Conversations



**TheOldMan&TheThree**  @OldManAndThree · Apr 29

On New York basketball. This is a special duo. Episode 55 with @hasanminhaj and @D\_Bo20 live now.



 12

 8

 71



**celcovington**  
@celcovington

Replying to @OldManAndThree @hasanminhaj and @D\_Bo20

@jj\_redick @talter Hello big fan of JJ and the pod here. Is there any reason why this ep is not uploaded to Spotify (just yet)? Thanks!

6:19 AM · Apr 30, 2021 · Twitter for Android

# Replies/Conversations



 **TheOldMan&TheThree** ✓ @OldManAndThree · Apr 29  
On New York basketball. This is a special duo. Episode 55 with @hasanminhaj and @D\_Bo20 live now.



12 8 71

 **celcovington** @celcovington · Apr 30  
@jj\_redick @talter Hello big fan of JJ and the pod here. Is there any reason why this ep is not uploaded to Spotify (just yet)? Thanks!

1


 **TheOldMan&TheThree** ✓ @OldManAndThree  
Replying to @celcovington @hasanminhaj and 3 others  
I think Spotify has been acting crazy all week

2:41 PM · Apr 30, 2021 · Twitter Web App



Saturday Night Live - SNL  @nbcสนl · May 23

Thank you [@anyataylorjoy](#), [@LilNasX](#), and [@chrisrock](#)!

And thank you to everyone who watched Season 46  [#SNLFinale](#)



 204

 491

 4.1K






# Hashtags



**Emma Watson** 

@EmmaWatson

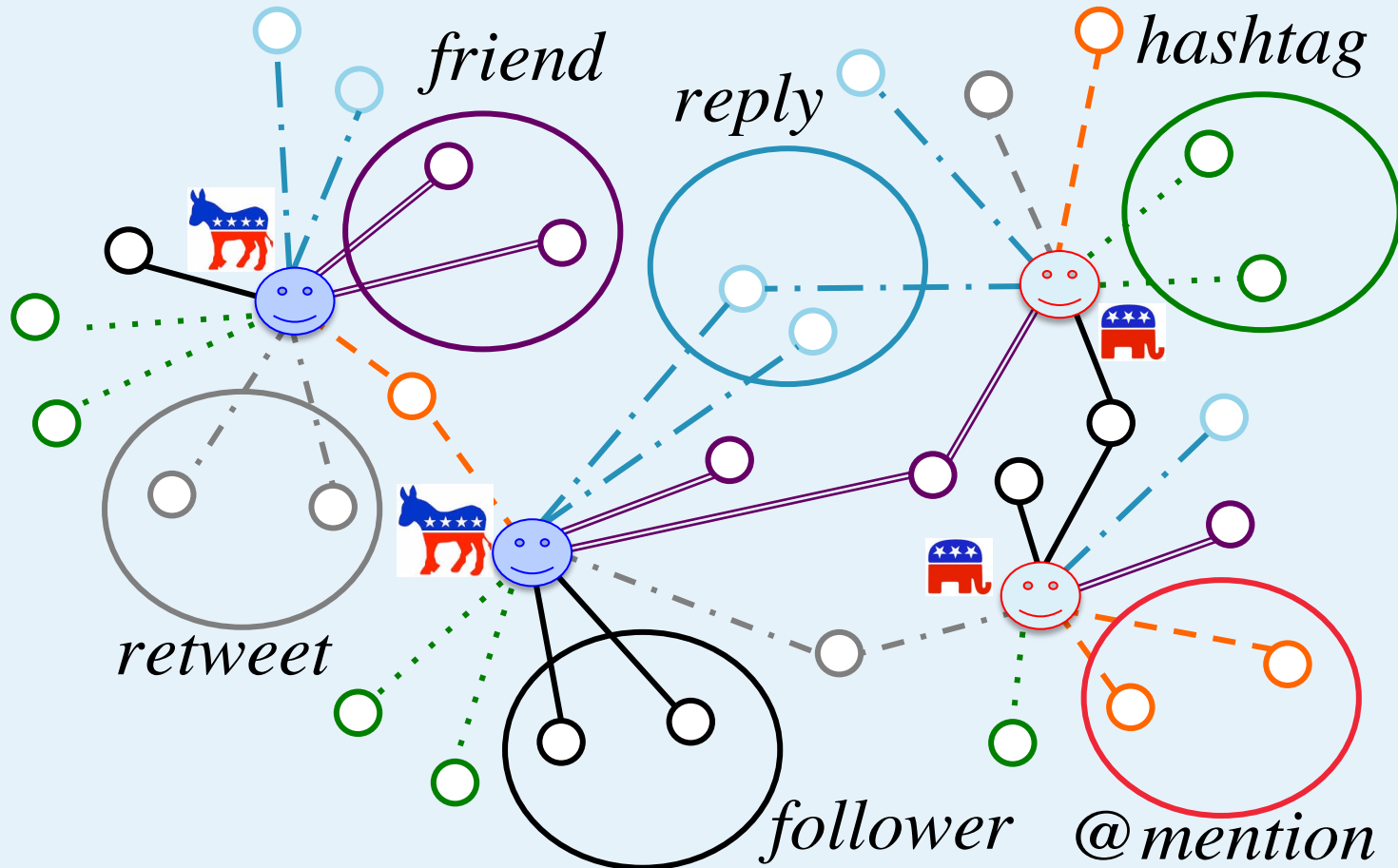


Gender equality not only liberates women but also men from prescribed gender stereotypes. [#heforshe](#) 

4:36 PM · Aug 18, 2014 · Twitter for iPhone

**21.9K** Retweets   **6** Quote Tweets   **28.3K** Likes

# Social Graph



Volkova, Van Durme, Yarowsky, Bachrach  
“Tutorial on Social Media Predictive Analytics” NAACL  
2015 [video slides](#)





# Twitter API



- One of the most open and powerful developer APIs of any major tech company
- Public API released in September 2006





## Streaming API

- Sample of public tweets in real time
- only real-time data
- Continuous connection
- No limited



## REST API

- Search, trends, author/user, interact
- Historical data up to one week
- One-time request
- Rate limited



## Streaming API

- Sample of public tweets in real time
- only real-time data
- Continuous connection
- No limited

## REST API

- Search, trends, author/user, interact
- Historical data up to one week
- One-time request
- Rate limited



- Way to provide authorized access to API

To start we need:

- A twitter account
- OAuth access tokens from <https://developer.twitter.com/en>





# Tweepy

**An easy-to-use Python library for accessing the Twitter API.**

 Fork 3,510  Star 7,772

Code: <https://github.com/tweepy/tweepy>

Documentation: <https://docs.tweepy.org/en/v3.10.0/>

# How to share Tweets



## Content redistribution

**The best place to get Twitter Content is directly from Twitter. Consequently, we restrict the redistribution of Twitter Content to third parties.** If you provide Twitter Content to third parties, including downloadable datasets or via an API, you may only distribute Tweet IDs, Direct Message IDs, and/or User IDs (except as described below). We also grant special permissions to academic researchers sharing Tweet IDs and User IDs for non-commercial research purposes.

In total, you may not distribute more than 1,500,000 Tweet IDs to any entity (inclusive of multiple individuals associated with a single entity) within any 30 day period unless you have received written permission from Twitter. In addition, all developers may provide up to 50,000 public Tweets Objects and/or User Objects to each person who uses your service on a daily basis if this is done via non-automated means (e.g., download of spreadsheets or PDFs).

Academic researchers are permitted to distribute an unlimited number of Tweet IDs and/or User IDs if they are doing so on behalf of an academic institution and for the sole purpose of non-commercial research. For example, you are permitted to share an unlimited number of Tweet IDs for the purpose of enabling peer review or validation of your research. If you have questions about whether your use case qualifies under this category please submit a request via the [API Policy Support form](#).

Any Twitter Content provided to third parties remains subject to this Policy, and those third parties must agree to the Twitter [Terms of Service](#), [Privacy Policy](#), [Developer Agreement](#), and [Developer Policy](#) before receiving such downloads. You may not enable any entity to circumvent any other limitations or restrictions on the distribution of Twitter Content as contained in this [Policy](#), the [Developer Agreement](#), or any other agreement with Twitter.

<https://developer.twitter.com/en/developer-terms/agreement-and-policy>