

Measurement and Collection of Social Network Data

Pawan Goyal

CSE, IITKGP

July 22, 2014

Social Network Site (SNS)

“A web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.” (Boyd and Ellison, 2007, p. 211)

What People Do in Social Media

- People make “friends” with others and build social relationships, connections and communities.
- People ask and answer one another.
- People create, publish or distribute information in the form of text, photos, video, audio, or tweets.
- People share bookmarks, presentation slides, or other files.
- People provide feedback on or rate others’ information.
- People create social tags or folksonomies.

Social Networking Websites

Name ↕	Description/Focus ↕	Date launched ↕	Registered users ↕	Registration
Google+	General	28 June 2011	1,600,000,000 ^[135]	Open to people 13 a ^[136]
Facebook	General: Photos, Videos, Blogs, Apps.	February 2004	1,280,000,000 ^[83]	Open to people 13 a
Qzone	General. In Simplified Chinese; caters for mainland China users		480,000,000 ^{[245][246]}	Open to the general
Sina Weibo	Social microblogging site in Mainland China.	14 August 2009	300,000,000 ^[260]	Open
Formspring	social Q&A website	November 2009	290,000,000 ^[100]	Registration post re questions
Habbo	General for teens. Over 31 communities worldwide. Chat room and user profiles.	August 2000	268,000,000 ^{[139][140][141]}	Open to people 13 a
Vkontakte	General, including music upload, listening and search. Popular in Russia and former Soviet republics.	September 2006	249,409,900 ^[307]	Open
LinkedIn	Business and professional networking	May 2003	200,000,000 ^[182]	Open to people 18 a
Renren	Significant site in China. Was known as 校内 (Xiaonei) until August 2009.		160,000,000 ^[250]	Open
Instagram	A photo and video sharing site	October 2010	150,000,000 ^[160]	Open to anyone 13 a
Bebo	General	July 2005	117,000,000 ^[16]	Open to people 13 a
Tagged	General.	October 2004	100,000,000 ^[281]	Open
Orkut	General. Owned by Google Inc. Popular in India and Brazil. ^[227]	22 January 2004	100,000,000 ^[228]	Open to people 18 a (Google login)
Netlog	General. Popular in Europe, Turkey, the Arab World and Canada's Québec province. Formerly known as Facebox and Redbox. ^[214]		95,000,000 ^[210]	Open to people 13 a
Twitter	General. Micro-blogging. RSS. updates	15 July 2006	93,800,000 ^[301]	Open to all ages ^[30]
Friendster	General. Popular in Southeast Asia. No longer popular in the western world	2002	90,000,000 ^[110]	Open to people 16 a
hi5	General. Popular in Nepal, Mongolia, Thailand, Romania, Jamaica, Central Africa, Portugal and Latin America. Not very popular in the USA.	2003	80,000,000 ^[143]	Open to people 13 a



[Larger Cover](#)

Mining the Social Web, 2nd Edition

Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More

By [Matthew A. Russell](#)

Publisher: O'Reilly Media

Final Release Date: October 2013

Pages: 448

★★★★☆ 4.6

[Read 11 Reviews](#)

[Write a Review](#)

How can you tap into the wealth of social web data to discover who's making connections with whom, what they're talking about, and where they're located? With this expanded and thoroughly revised edition, you'll learn how to acquire, analyze, and summarize data from all corners of the social web, including Facebook, Twitter,...

[Full description](#)

IPython Notebook

May use the following url: `http://smash.psych.nyu.edu/courses/spring12/modeling/ipythonhints.html`

A microblogging service, allowing people to communicate with short, 140-character messages that roughly correspond to thoughts or ideas.

A microblogging service, allowing people to communicate with short, 140-character messages that roughly correspond to thoughts or ideas.

Twitter's relationship model

A microblogging service, allowing people to communicate with short, 140-character messages that roughly correspond to thoughts or ideas.

Twitter's relationship model

Allows you to keep up with the latest happenings of any other user, even though the other user may not choose to follow you back or even know that you exist.

A microblogging service, allowing people to communicate with short, 140-character messages that roughly correspond to thoughts or ideas.

Twitter's relationship model

Allows you to keep up with the latest happenings of any other user, even though the other user may not choose to follow you back or even know that you exist.

Interest Graphs

A way of modeling connections between people and their arbitrary interests.

Fundamental Twitter Terminology

Tweets are the essence of Twitter. In addition to the textual content, we get:

Tweet entities

User mentions, hashtags, URLs and media

Places

Locations in the real world, attached to a tweet

@ptwobrussell is writing @SocialWebMining, 2nd Ed. from his home office in Franklin, TN. Be #social: <http://on.fb.me/16WJAf9>

Tweet entities

- User mentions: @ptwobrussell, @SocialWebMining
- hashtag: #social
- URL: <http://on.fb.me/16WJAf9>
- Place: Franklin, Tennessee, in the tweet and a location metadata, where the tweet is authored

Timelines

Chronologically sorted collections of tweets

- **home timeline:** view that you see when you log into your account and look at all the tweets from users that you are following:
`https://twitter.com`
- **user timeline:** collection of tweets only from a certain user:
`https://twitter.com/timoreilly`
- **User's home timeline:** can be accessed with the additional *following* suffix, appended to the URL. :
`https://twitter.com/timoreilly/following`



Tim O'Reilly

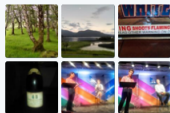
@timoreilly

Founder and CEO, O'Reilly Media. Watching the alpha geeks, sharing their stories, helping the future unfold.

Sebastopol, CA
radar.oreilly.com
Joined March 2007

Tweet to Tim O'Reilly

186 Photos and videos



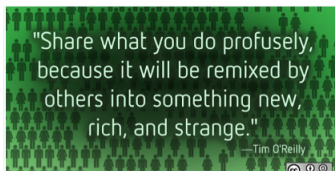
TWEETS 30.3K PHOTOS/VIDEOS 186 FOLLOWING 1,222 FOLLOWERS 1.78M FAVORITES 274 More

Tweets Tweets and replies

Retweeted by Tim O'Reilly

Open Source Way @opensourceway · 6h

In honor of #OSCON, we had to share this quote from @timoreilly that captures the #opensource way perfectly.



42 19 View more photos and videos

Retweeted by Tim O'Reilly

DeWitt Clinton @dewitt · 6h

@timoreilly @timbray And I'd argue Go is catching on precisely because the number of things you need to know to do it right is so small.

13 7 View conversation

Tim O'Reilly @timoreilly · 6h

IndieWebCamp looks awesome bit.ly/1u8wABa #indieweb via

user home timeline



Tim O'Reilly ✓

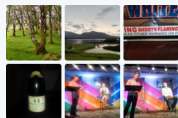
@timoreilly

Founder and CEO, O'Reilly Media. Watching the alpha geeks, sharing their stories, helping the future unfold.

📍 Sebastopol, CA
🌐 radar.oreilly.com
🕒 Joined March 2007

Tweet to Tim O'Reilly

📷 186 Photos and videos



TWEETS **30.3K** PHOTOS/VIDEOS **186** FOLLOWING **1,222** FOLLOWERS **1.78M** FAVORITES **274** More ▾

⚙️ Follow

Bitcoin Today @NewBitcoinToday
If you have to ask what bitcoin is, you'll never know

Robot Film Festival @robotfilmfest
4th Annual RFF will be July 19 at @botndolly in SFI Red carpet! Robot films! Botskers Award Ceremony! @heatherknight @beat_bots @macdiva

Agnes Crawford @understandrome
Architectural historian & qualified Rome guide. 12 years' experience giving tours in the city. Best job in the world! understandingrome.wordpress.com

congress-edits @congressedits
I'm a bot that tweets anonymous Wikipedia edits that are made from IP addresses in the US Congress. You can

Walter Isaacson ✓ @WalterIsaacson
CEO of the Aspen Institute. Author of The Innovators (Oct 2014) and biographies of Jobs, Einstein, Franklin,

Fabian Giesen @rygorous
hund.



TweetDeck: a highly customizable user interface

The screenshot displays the TweetDeck interface with several columns:

- New Tweet:** A blue header with a plus icon and the text "New Tweet".
- Search:** A search bar with the placeholder text "Search".
- Interactions:** A column with a list of interaction types: Interactions (@pwbcrussell), Interactions (@SocialWebMining), Activity (@pwbcrussell), Me (@SocialWebMining), and Timeline (@SocialWebMining). Below this is an "Add column" button.
- Interactions @SocialWebMining:** A column with a "SHOWING" dropdown set to "All interactions". Below it are sections for "CONTENT", "USERS", "ALERTS", and "PREVIEWS". A "Clear column" button is at the bottom.
- Activity @pwbcrussell:** A column showing a list of tweets, including one from DeWitt Clinton favorited Om Malik.
- @ Me:** A column with a "Filtered by:" dropdown and a list of tweets, including one from Paul Ivanov.
- Timeline @SocialWebMining:** A column showing a list of tweets, including one from MiningTheSocialWeb retweeted O'Reilly Strata.

Samples of public tweets flowing through Twitter in realtime

Public firehose

Known to peak at hundreds of thousands of tweets per minute during events with particularly wide interest.

Public timeline

A small random sample of the public timeline is available, that provides filterable access to enough public data for API developers

Creating Twitter API connection

Create an application at <https://dev.twitter.com/apps>.

 Application Management



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.



Create a new Twitter application to get OAuth credentials and API access

- Consumer key
- Consumer secret
- access token
- access token secret

https:

`//github.com/ptwobrussell/Mining-the-Social-Web-2nd-Edition`

Requires : IPython Notebook

Things to try out

- Exploring Trending Topics
- Searching for tweets using a query
- Analyzing the 140 characters

Computing Lexical Diversity

Number of unique tokens in the text, divided by the total number of tokens in the text

Can be a primitive statistic for

- How broad or narrow the subject matter is that an individual or group discusses
- Comparing different groups/individuals
- Comparing across time periods

Accessing Facebook user's account data

Register an application as the entry point into the Facebook developer platform.

Data accessible

Whatever the user has explicitly authorized it to access.

As a developer, the application would be like any of your Facebook friends, in that you are ultimately in control of what the application can access.

Accessing Facebook user's account data

Register an application as the entry point into the Facebook developer platform.

Data accessible

Whatever the user has explicitly authorized it to access.

As a developer, the application would be like any of your Facebook friends, in that you are ultimately in control of what the application can access.

Developer Tools

Graph API Explorer app for querying the Social Graph

Accessing Facebook user's account data

Register an application as the entry point into the Facebook developer platform.

Data accessible


Whatever the user has explicitly authorized it to access.

As a developer, the application would be like any of your Facebook friends, in that you are ultimately in control of what the application can access.

Developer Tools

Graph API Explorer app for querying the Social Graph

You can translate your queries into Python code for automation

facebook developers [Docs](#) [Tools](#) [Support](#) [News](#) [Apps](#)  Matthew Russell

Graph API Explorer

Home > Tools > Graph API Explorer Application: [?] **Graph API Explorer**

Access Token: [Debug](#) [Get Access Token](#)

Graph API [FQL Query](#)


GET

[Learn more about new features.](#)

Node: 644382747

- id
- name
- +

```
{
  "id": "644382747",
  "name": "Matthew A. Russell"
}
```

facebook developers Docs Tools Support News Apps  Matthew

Graph API Explorer

Home > Tools > Graph API Explorer Application: [?] **Graph API E**

Access Token: Debug Get Acc

Graph API FQL Query

→ / 644382747?fields=id,name,friends

Learn more about new features.

Node: 644382747

- id
- name
- friends →
 - + _____
 - + _____

```
{
  "id": "644382747",
  "name": "Matthew A. Russell",
  "friends": {
    "data": [
      {
        "name": "Bas Russell",
        "id": "6224364"
      },
      {
        "name": "Derek Brown",
        "id": "51801220"
      }
    ]
  }
}
```

Graph API query

facebook developers

Search Facebook Developers



Docs

Tools

Support

News

Apps



Matthew A. Russell

Graph API Explorer

Home > Tools > Graph API Explorer

Application: [?] **Graph API Explorer**

Access Token:

Graph API FQL Query

→

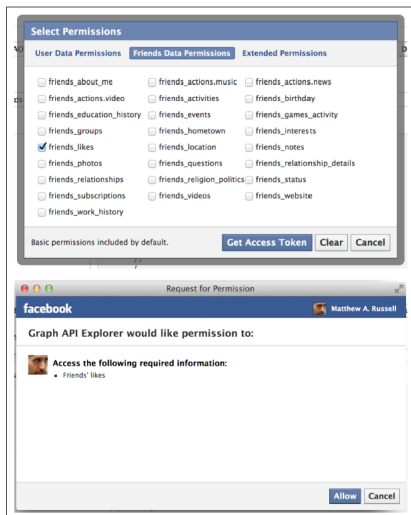
Learn more about new features.

Node: 644382747

- id
- name
- friends →
 - likes
 - + _____
 - + _____
 - + _____

```
{
  "id": "644382747",
  "name": "Matthew A. Russell",
  "friends": {
    "data": [
      {
        "id": "6224364"
      },
      {
        "id": "51801220",
        "likes": {
          "data": [
            {
              "category": "Sports/recreation/activities",
              "category_list": [
                {
                  "id": "184405378265823",
                  "name": "Gym"
                }
              ]
            },
            {
              "name": "SOP WODs",
              "id": "226640360698227",
              "created_time": "2013-04-03T17:50:41+0000"
            }
          ]
        }
      },
      {
        "category": "Local business",
        "category_list": [
```

Graph API query



Analyzing Facebook Pages

```
# Get an instance of Mining the Social Web
# Using the page name also works if you know it.
# e.g. 'MiningTheSocialWeb' or 'CrossFit'
mtsw_id = '146803958708175'
pp(g.get_object(mtsw_id))
```

Sample output for the query reveals the data that backs the object's Facebook page, as shown here:

```
{
  "category": "Book",
  "username": "MiningTheSocialWeb",
  "about": "Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social...",
  "talking_about_count": 22,
  "description": "Facebook, Twitter, and LinkedIn generate a tremendous ...",
  "company_overview": "Like It here on Facebook!\n\nFollow @SocialWebMining...",
  "release_date": "January 2011",
  "can_post": true,
  "cover": {
    "source": "https://sphotos-b.xx.fbcdn.net/...",
    "cover_id": 474206292634605,
    "offset_x": -41,
    "offset_y": 0
  },
  "mission": "Teaches you how to...\n\n* Get a straightforward synopsis of ...",
  "name": "Mining the Social Web",
  "founded": "January 2011",
  "website": "http://amzn.to/d1Ci8A",
  "link": "http://www.facebook.com/MiningTheSocialWeb",
  "likes": 911,
  "were_here_count": 0,
  "general_info": "Analyzing Data from Facebook, Twitter, LinkedIn, ...",
  "id": "146803958708175",
  "is_published": true
}
```

Further Exploration

Things to try out

- Analyzing things your friends “like”: top ten things, categories
- Common likes between an ego and its friendship