CSE 472: Social Media Mining

Project I - Social Media Data Analysis

Prof. Huan Liu
TA: Ujun Jeong (ujeong1@asu.edu)
Due at 2021, September 28th, 11:59PM

This is an ***individual*** project assignment. Before you start the project, you are required to fill out a survey on your desired choice of social media platform (See the "Survey Guidelines" section for details). Once the platform has been assigned, you may begin working on the steps in the project guideline and create the report, dataset, and codes (See the "Submission" section for details).

Project Objectives

In this project, you will learn how to crawl social media data as well as process and perform exploratory analysis on your extracted data.

Survey Guideline

Go to the "Quizzes" tab in Canvas and fill out the survey for Project 1. Choose one of the social media platforms in the **first week** of your project among (1) Twitter, (2) Facebook, (3) Telegram, (4) Reddit, and (5) Other. We strongly recommend using an API for crawling data to avoid any possible violation of data policy. If you are assigned option 5, you need to be sure that there is an API that can be used.

We will limit the number of students per category to make it distributed equally (i.e., the total number of students will be divided by five). The assignment of the platforms will be first come first served by checking out the timestamp of your submitted survey. If the number of students in a category exceeds its limit, we will consider your other priorities or randomly assign you to one of the five categories.

Project Guideline

Step 1. Obtain the credentials for scraping the data (i.e., API-key). Detailed instructions for obtaining the credentials for some of the platforms are described in the appendix. Some APIs may take time to obtain credentials. Therefore, we recommend requesting credentials as fast as you can.

After you get the credentials, you need to learn how to use API by yourself. You can refer to the official documentation or utilize existing frameworks and libraries such as SDK or wrappers as follows:

1. Twitter: Twark, Twint, Tweepy
2. Facebook: Python SDK
3. Telegram: Telethon

4. Reddit: PRAW

Save the data you collected in the JSON format and write a brief description in your report on how you crawled your data.

Step 2. Now that you have fetched the data, construct and visualize your data as a graph. There are multiple packages and softwares available for graph analysis such as, networkx, snap, Gephi, NodeXL and graph-tool. Choose one and read the instruction on how to build your graph. For each package, it may require a certain format of the graph data: edgelist, adjacency matrix or adjacency list.

In this step, you are asked to create a social network with at least around 100-500 nodes. Some representative network types are described as follows, but you are free to define your own network type as long as you can justify it. In the report, write what features you used for constructing the graph and take a snapshot of your returned graph and add it to your report.

1. Friendship Network. A user's friendship network can be represented as a graph where the nodes are the users and the edges show whether there is a friendship relationship between them. Example: Users and their follower and followee relationship as a directed graph.
2. Diffusion Network. A node represents a user, which can publish, receive, and propagate information. A directed edge between nodes represents the direction of information propagation. Example: news propagation where the nodes are users and the edges are reposts.
3. Word Co-occurrence Network. A network in which the nodes are the words and two words occurring together are linked by an edge.
4. Co-authorship Network. The nodes are scientists, and two scientists are connected if they have coauthored a paper. Example: an authorship network in "Computer Science" category papers.

Step 3. You will learn different network measures in class (Degree Distribution, Clustering Coefficient, Pagerank, Diameter, Closeness, Betweeness, etc.). Use your chosen package or software from step 2 to obtain "Degree Distribution" and plot it as *histogram*. Beside this measure, choose two other measures from what you have learned and plot them in case they are returned as distribution, or a number, otherwise. Attach your results to your report.

Submission

We will run your code to see if it works for all the steps. You should put all your files including your data, source code, and a report (preferred to be PDF format). Submit the report and related files on either "Gradescope" or "Assignment" tab in Canvas. You can decide whether to upload your submission as zip file or not because it will be automatically uncompressed once you upload the zip file.

Most importantly, you are only allowed to use **Python** as the programming language. Please make sure your source code has the **".py"** extension. i.e., If you used iPython, then you have to convert it into ".py" extension before you submit your codes.

In summary, your submission should contain the following items:

1. The project report that satisfies the requirements suggested in each step.
2. The dataset generated by your code (The dataset should be in the format of ".json" file).
3. Source codes with ".py" extension and their related files (e.g., config file, text file, and etc).

Grading Criteria

pts	Description
1	Select platform in the first week
3	Step 1: Data Collection
3	Step 2: Graph Construction & Visualization
3	Step 3: Network Measures Calculation
10	

Table 1: Grading Rubric

Academic Integrity

- To prevent any potential plagiarism, we will randomly select students for each phase of the project and ask them to talk with the TA.
- Your codes in the submission will be automatically checked by the similarity detection tools.
- For Step 1, you have to develop your own code for data scraping. It is NOT permissible to use a public available datasets.
- For all the three steps, you can *refer* to others' code and use libraries, software and packages but it is NOT permissible to copy existing code from others.
- Use a "Reference" section and cite all the tutorials, packages, software and libraries you used in your data.

APPENDIX: Instructions to obtain API-keys

Twitter:

1. Visit https://developer.twitter.com
2. Log in to your Twitter account or Sign up for a new one.
3. In the top right hand corner click "Apply". Then click "Apply for a developer account".
4. For your primary reason for using Twitter developer tools choose "Student". Then click "Next".
5. You must add a valid phone number to your account to use the Twitter API. Then add your country as "United States" and pick some name for the developer account (you may use your email as a username).
6. Complete the form on how you intend to use your Twitter Developer Account. We recommend rewording the following answers:
   (a) *In your words (how you plan to use Twitter data and/or APIs...)* "I am taking the CSE 472 Social Media Mining course at Arizona State University under Dr. Huan Liu. For my class project, I will use the Twitter API to access user network information and analyze the data using standard metrics."
   (b) *Are you planning to analyze Twitter data?* **Yes** "I will be calculating network measures such as Degree Distribution, Clustering Coefficient, Pagerank, Diameter, Closeness, Betweeness, etc. for my project."
   (c) *Will your app use Tweet, Retweet, like, follow, or Direct Message functionality?* **No**
   (d) *Do you plan to display Tweets or aggregate data about twitter content outside of Twitter?* **Yes** "I will be displaying users as nodes in a graph and use follow / friend relationships as edges. The output will only be displayed in class during the project presentation."
   (e) *Will your product, service, or analysis make Twitter content or derived information available to a government entity?* **No**
7. Review your previous answers and accept the terms and conditions. Then wait for your developer account to be approved. This could take up to a couple days.
8. Once your account has been approved, log back into https://developer.twitter.com and in the top right hand corner click there will be a dropdown menu just to the left of your profile photo. Choose "Apps".
9. Click "Create an app".
10. Fill out the four required fields (App name, Application Description, Website URL, and Tell us how this app will be used) then click "Create".
11. Once the app has been created, you should be able to click "Details" then "Permissions". Configure your application to be Read-only.
12. Next go to the "Keys and tokens" tab. You will see both **Consumer API keys** and **Access token & access token secret**. Use these to invoke Twitter API calls.

Facebook:

1. Visit https://developers.facebook.com
2. Log in to your Facebook account or Create a New Account.
3. In the top right hand corner click "Get Started".
4. You must then verify your phone number to your account to use the Facebook Developer API.
5. Fill out the "about you" section (reference #6 on the Twitter instructions for some example answers to these questions).
6. Review your previous answers and accept the terms and conditions. Then wait for your developer account to be approved. This could take up to a couple days.
7. Once your account has been approved, log back into https://developers.facebook.com and in the top right hand corner there will be a dropdown called "My Apps". Choose "Create App".
8. Fill out the two required fields (Display Name and Contact Email) then click "Create App ID".
9. Complete the reCAPTCHA security check.
10. On the following "Add a Product page", you can choose to add a product such as the "Facebook Login". Once you have selected a product then it will appear on the left hand menu under "Products".
11. Under "Settings" choose "Basic". There you will find your **App Id** as well as your **App Secret** which will be used to access the Facebook API later.
12. Now to get an access token, under "Roles" choose "Test User". Click edit on the default user and choose "Get an access token for this user". Copy the string to your script. Now that you have an **Access Token**, you are ready to use the Facebook API.
13. To make development even easier, you can create a "Test App" by clicking on the dropdown that has your App's name at the top left and choosing "Create Test App". Having a test app will allow you to use unencrypted web addresses for Facebook Login among other things.
14. Depending on the tools you want to use, it will require further permissions. You can find how to request permissions on the App Dashboard. For more details for the dashboard, refer to the official site https://developers.facebook.com/docs/development/create-an-app/app-dashboard/.
15. Note that requesting advanced permissions might require verification of your application which needs well organized documentation. We recommend to use some basic permissions and permissions that are already granted to your account because requestments take time to get accepted.

Telegram:

1. For the Telegram API, you need to have a phone number available.
2. In case you do not want to use your own home/work phone numbers, you can create a virtual one with apps that will provide you some, like TextNow.
3. However, Telegram can ban your account once they detect you are using the temporarily phone number. So the decision is totally up to you.
4. Install the Telegram app and use your phone number to connect. Telegram will send you a verification code. For a free virtual number, since it is a restricted account, you might not get the verification code. This feature is usually provided by upgraded accounts. But if you wait for around 2-3 minutes, Telegram will call your phone and provide you with the code.
5. Visit Telegram's API development tools page.
6. Enter your phone number.
7. A confirmation code will be messaged to you in the Telegram app.
8. In the "Create New Application" page, enter the "App Title", a "Short Name" for your app, a "URL" (e.g. http://127.0.0.1/), the "Platform" (preferably Desktop), and a short "Description" of your app (e.g. This is a test app for Project I of social media mining course).
9. You will be provided with app configuration including **api_id** and **api_hash**. Save these credential tokens and use them in your program.

Reddit:

1. First of all, sign into Reddit homepage and check your **USERNAME** and **PASSWORD** to proceed the steps (USERNAME is not same as your account if you sign in by Google or Apple)
2. Visit https://www.reddit.com/prefs/apps to get tokens required for using API.
3. Click "create another app..." and fill out the required details. Make sure to select script as an option. Then, you can create an application by clicking "create app".
4. Click "edit" on your app, and save your keys belong to **"personal use script"** and **"secret"**.
5. Request a temporary OAuth token from Reddit by following this python code. Finally, you get an access token which lasts about 1 hour.
6. If you are interested in this python code, you can refer to the original guideline in this video