

CS 579: Online Social Network Analysis

Project I - Social Media Data Analysis

Prof. Kai Shu

Due on Feb. 14, 2022 at 11:59 pm

This is a *2-member* project assignment. Each group is supposed to work on the steps together, including the writeup of the report. Please submit one set of your report and related files per group on Blackboard.

Project Objectives

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

Project Outline

1. Data Collection
2. Data Visualization
3. Network Measures Calculation

Guideline

Step 1. Choose one of the social media platforms in the first week of your project such as Twitter, Facebook, Instagram, dblp, Reddit, arXiv, ResearchGate, Stackoverflow, Stackexchange and etc. Figure out how you can crawl data from these websites. Some of these websites provide API for crawling the data. Make sure you have the needed credentials for scraping the data (i.e. API-key).

You will be asked to create the social network with at least around 100-500 nodes. Some representative networks types are described as follows,

1. Friendship Network. A user's friendship network can be represented as a graph that the nodes are the users and the edges shows whether there is a friendship relation between them. Example: Users and their follower and followee relationship in Twitter as a directed graph.
2. 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 in arXiv website.
3. Diffusion Network. A node represents an entity, which can publish, receive, and propagate information. A directed edge between nodes represents the direction of information propagation. Example: fake news propagation when the nodes are users and the edges are re-tweets/replies/likes.

Write a brief description in your report on how you crawled your data.

Step 2. Now that you fetched the data, visualize your network as graph. There are multiple packages and softwares available for graph analysis such as, networkx [\[link\]](#), snap [\[link\]](#), Gephi [\[link\]](#), NodeXL [\[link\]](#) and graph-tool [\[link\]](#). Choose one and read the instruction on how to visualize your graph (for each package, it may require a certain format of the graph data: edgelist, adjacency matrix or adjacency list). Take a snapshot of your returned graph and add it to your report.

Step 3. You will learn different network measures in class (Degree Distribution, Clustering Coefficient, Pagerank, Diameter, Closeness, Betweenness, 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 codes to see if it works for all the steps. You should put all your files including your data, source codes, pdf file (report) into a *.zip* folder named LASTNAME1_LASTNAME2_PJ1 (Instead of LASTNAME1 and LASTNAME2 type the lastname of each member). Submit your zip folder on Blackboard.

pts	Description
1	Select platform in the first week
3	Data Collection
3	Data Visualization
3	Network Measures Calculation
10	

Table 1: Grading Rubric

Academic Integrity

- For step 1, you have to develop your own code for data scraping. It is NOT okay to use public available dataset.
- For all the three steps, you can *refer* to others' code and use libraries, software and packages but it is NOT okay to copy existing code from others.
- Use a "Reference" section and cite all the tutorials, packages, software and libraries you used in your data.

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. On 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 intent 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 CS 579 Online Social Network Analysis course at Illinois Institute of Technology under Dr. Kai Shu. 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, Betweenness, 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. 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.