Enter your responses inline below and push this file to your private GitHub repository.

1. Assume I plan to use the friend\_overlap function above to quantify the similarity of two users. E.g., because 344 is larger than 1, I conclude that

Jill Stein and Gary Johnson are more similar than Hillary Clinton and Donald Trump. How is this approach misleading? How might you fix it?

* It is true for Jill Stein and Gary Johnson that they have more friends than Hillary Clinton and Donald Trump. As it is seen by the function count\_friends which gives following output:

DrJillStein 4136

GovGaryJohnson 3052

HillaryClinton 745

realDonaldTrump 42

So if users, Jill Stein and Gary Johnson, have more number of friends than, Hillary Clinton and Donald Trump, they are bound to have more similar or common friends. Whereas for Hillary Clinton and Donald Trump, they have less number of friends so they are bound to have less similar or common friends. This approach is misleading as we are not considering the total number of friends of each candidate into consideration, which is a very important factor to find out similarity between two candidates.

To fix this approach I would use Jaccard coefficient which is the ratio of common friends of both candidates to common plus uncommon friends of both candidates.

This Jaccard coefficient will give the real similarity between two candidates.

2. Looking at the output of your followed\_by\_hillary\_and\_donald function, why do you think this user is followed by both Hilary Clinton and Donald Trump, who are rivals? Do some web searches to see if you can find out more information?

* Hillary Clinton follows Earvin Magic Johnson because he is Democratic party supporter and Hillary’s support for the Presidential election 2016. In the past when Hillary ran for president in 2007, he supported that time too. Hillary and Earvin Magic Johnson are friends and he has hosted a fundraiser for Hillary Clinton's presidential campaign on August 22, 2016.

Earvin Magic Johnson is followed by Donald Trump as he has been a basketball player and is a frequent guest at Donald Trump’s hotel. He’s a businessman too just like Donald Trump. He is open to doing business with Donald Trump and consider him a friend but a very staunch Hillary supporter.

So, Earvin Magic Johnson is a followed by Donald Trump due to his professional similarity and by Hillary Clinton as he is her supporter and Media figure for fund raising.

3. There is a big difference in how many accounts each candidate follows (Gary Johnson follows over 3K accounts, while Donald Trump follows less than 50). Why do you think this is? How might that affect our analysis?

* The candidates like 'DrJillStein', 'GovGaryJohnson' follow more candidates as they are not very famous and following more people and tweeting regularly will make then appear top on feed list so more people will hear their views and be interested in their presidential campaign.
* As for Donald Trump and Hillary Clinton, they are well known presidential candidates and do not need much following. They are on top list of winning the presidential post so they do not need to follow much people except for close friends and campaign supporters. They are the ones who are googled the most and they get publicity more than the other two.
* The number of friends being followed should not affect the analysis if a correct method is followed for analysis.

4. The follower graph we've collected is incomplete. To expand it, we would have to also collect the list of accounts followed by each of the friends. That is, for each user X that Donald Trump follows, we would have to also collect all the users that X follows. Assuming we again use the API call https://dev.twitter.com/rest/reference/get/friends/ids, how many requests will we have to make? Given how Twitter does rate limiting (https://dev.twitter.com/rest/public/rate-limiting), approximately how many minutes will it take to collect this data?