Homework 12: Large Retweet Networks

Started: Dec 4 at 9:55pm

Quiz Instructions

Now that we have some experience harvesting and analyzing Twitter networks, it's time to apply those skills to a larger network, involving the sort of real-world issues we've discussed in class. The questions involving calculations which have been covered in previous assignments are weighted less heavily this time - the emphasis is on your analysis, so keep that in mind when writing your answers.

This assignment deals with a large retweet network from a topic which trended a few years ago. To ensure everyone is working with the same data - use the already-harvested file of tweets containing the hashtag #RepealThe19th (**background info** ☑ **(http://www.snopes.com/2016/10/13/donald-trump-supporters-repealthe19th/)**). **The file is named** `repealthe19th.jsonl.gz` and it includes 23343 tweets. After parsing the file, you should verify that you have this many tweets; a deviation from this number indicates you had an issue parsing and it may affect any or all of your answers.

- Do all of your work in a single notebook, label your answers, and attach your completed notebook to your submission.
- Make sure the notebook runs top-to-bottom by clicking "Restart & Run All" from the "Kernel" menu.
- **I may spot-check your notebook**, so make sure you clearly label notebook output corresponding to exercise answers.
- You will receive zero credit for submitted answers that don't match the work in this notebook!
- Rename your file lastname_repealthe19th_retweet before submitting

---

Question 1	0.5 pts

What proportion of tweets in this list are retweets? Round to three decimal places.

```

```

Question 2	0.5 pts

As we did in the tutorial, I want you to create a retweet network in order to answer the following questions. Things to remember:

1. This should be a directed graph. Nodes are users, and edges represent retweets.

2. Edges go in the direction of information flow: if you retweet me, there is an edge from me to you.
3. **YOU SHOULD REMOVE SELF-LOOPS**. You can remove them after creating the graph, or you can modify your graph creation code to not add them at all. This is in bold because it will affect some answers.

How many nodes are in the retweet network?

Question 3   0.5 pts

How many edges are in the retweet network?

Question 4   0.5 pts

What is the highest out-degree in the retweet network? I want the out-degree here, not the screen name (that comes later).

Question 5   0.5 pts

What is the highest in-degree in the retweet network? I want the in-degree here, not the user.

Question 6
0.5 pts

Which node has the highest out-degree in the network? Answer with the user's screen name.

Question 7
0.5 pts

Which node has the second-highest out-degree in the network? Answer with the user's screen name.

Question 8
0.5 pts

What proportion of nodes in this network have zero out-degree? Round to three decimal places.

Question 9
1 pts

In your opinion, what could this proportion indicate about this particular network?

HTML Editor

B I U A ▾ A ▾ I_x ≡ ≡ ≡ ≡ ≡ x^2 x_2 ☰ ☰

⊞ ▾ ▣ 🔗 ✂ 🖼 √x ✳ ≈ V ▶ ¶ ¶ 12pt ▾ Paragraph

Question 10   0.5 pts

Which of the following best describes what it means for a node to have zero out-degree in this network? Each of the statements applies only to the sample of tweets we used to make this network.

○ The user produced no tweets

○ The user hasn't retweeted anyone else

○ The user hasn't been retweeted by anyone

○ The user has no followers

○ The user does not follow any other users

Question 11   0.5 pts

Which of the following best describes the connectivity of this network?

○ Strongly connected

○ Weakly connected

○ Connected

○ Unconnected

Question 12

0.5 pts

What proportion of this network's nodes are in its largest weakly-connected component? Round to three decimal places.

Question 13

1 pts

In your opinion, what does could the connectivity of this network imply?

HTML Editor

B I U A ▾ A ▾ I_x ≡ ≡ ≡ ≡ ≡ x² x₂ ☰ ☰

⊞ ▾ ▣ 🔗 ⤫ 🖼 √x ✳ 〽 V ▷ ¶ ¶ 12pt ▾ Paragraph

0 words

Question 14   1 pts

Use your list of tweets and what you know about retweets to do the following problem.

Your task is to find the most retweeted tweet. To do so, you should count the number of times each tweet id is retweeted in your list of tweets. A dictionary is probably useful for this task, with the key as the tweet id and the value as the number of times it's been retweeted. You could also use `collections.Counter` if you like. Note that each tweet's id is found under the `'id_str'` key.

Once you have the id of the most retweeted tweet, you can see the tweet by visiting the following URL in your browser

```
https://twitter.com/statuses/{id}
```

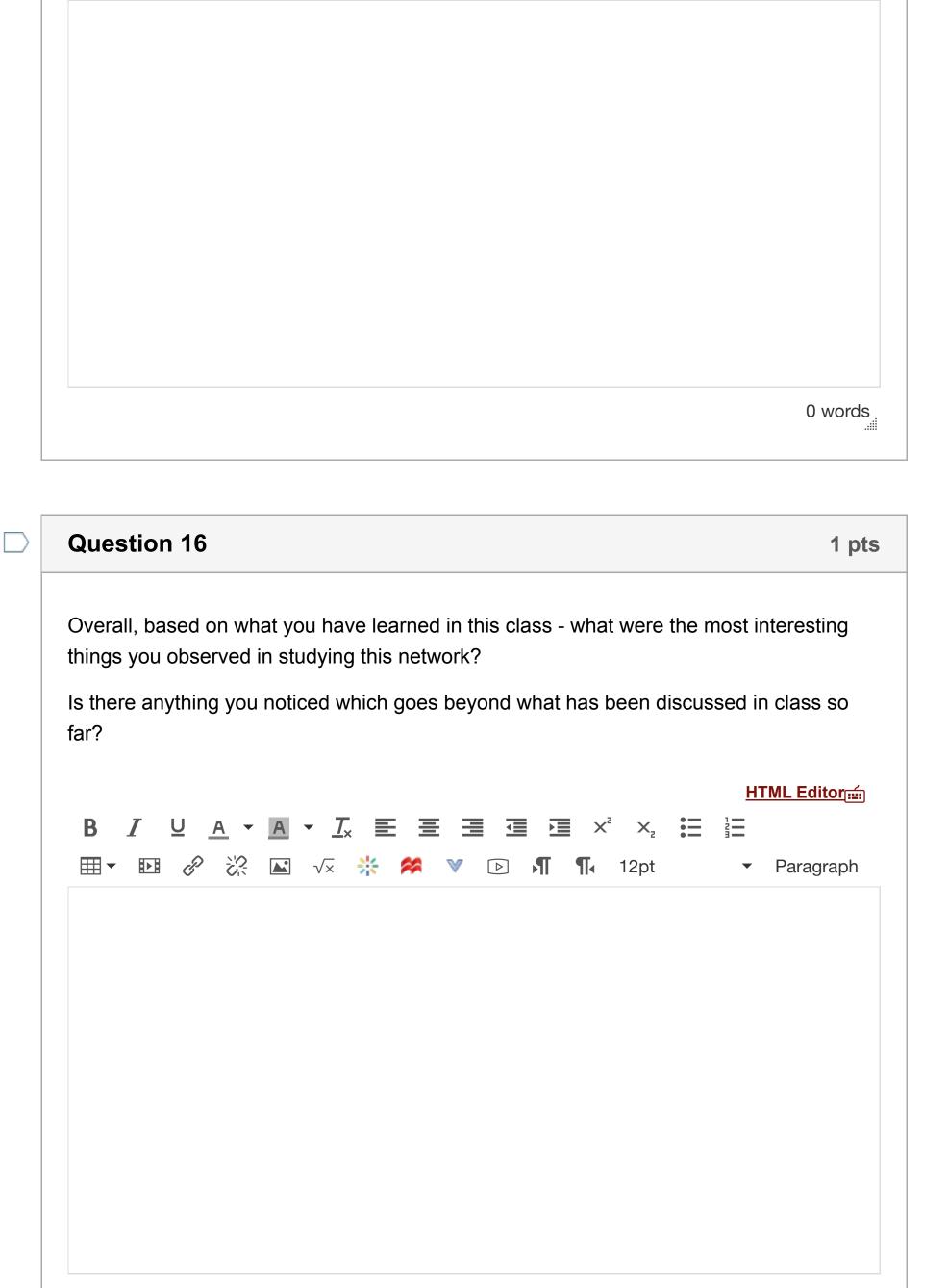replacing everything after the last slash with the actual tweet ID.

Since Canvas is not good at copy/paste with complex text, just submit the tweet ID as the answer to this problem.

Question 15   1 pts

Do you think there are any signs of astroturfing? (bots, coordinated campaigns, etc?) Justify your response.

HTML Editor

B *I* U A ▾ A ▾ I̵ₓ ☰ ☰ ☰ ☰ ☰ x² x₂ ☰ ☰

⊞ ▾ ▣ 🔗 ⌇ 🖼 √x ✳ 〰 ∨ ▣ ¶ ¶ 12pt ▾ Paragraph

0 words

Question 16

1 pts

Overall, based on what you have learned in this class - what were the most interesting things you observed in studying this network?

Is there anything you noticed which goes beyond what has been discussed in class so far?

B *I* <u>U</u> A ▾ A ▾ T̲ₓ ≡ ≡ ≡ ≡ ≡ x² x₂ ☰ ☰

⊞ ▾ ▦ 🔗 ⚯ 🖼 √x ✳ 〰 V ▶ ¶ ¶ 12pt ▾ Paragraph

Question 17

0 pts

Upload your notebook as the solution for this question.

Remember to **label your answers** to the exercises so I can find them when spot-checking your submitted answers, and make sure the notebook runs top-to-bottom by clicking "Restart & Run All" from the "Kernel" menu. You will receive zero credit for submitted answers that don't match the work in this notebook!

Upload   Choose a File

Quiz saved at 4:01pm   Submit Quiz