Assigments Description

s232892 Zeng

November 2024

1 Feedback from class

1. The information of the network is big enough however we need to pay attention to the domination point(super high degree like luffy), we have many of them and these node can't be partition into community rightly since in each community they have equal possibility, meanwhile the content of this node will interference the analysis of NLP of the community.
2. we can have more information on other platforms like reddit
3. we should introduce the time dimension to explore the changes of the network or people's relationship this is the main point of storyline analysis.
4. When we do the sentimental analysis the word-list maybe not enough since the animation contain many words very special.

2 Main Task

1. Explore web APIs for data collection.
2. Apply a high level programming language (e.g. Python) to utilize such APIs for data acquisition.
3. Apply natural language processing to represent statistical structures in text and analyze the content.
4. Apply and discuss the main strategies for detecting sentiment in media (e.g. text, music, images, etc).
5. Apply standard algorithms to recommend media (text, audio, video) according to user preferences and user context (friends, mood, location, etc).
6. Assess basic metrics for complex networks, and model social relations based on network analysis.
7. Implement software for detecting communities in social networks and analyze the communities using network metrics.

8. Quantify relations in social networks to analyze their dynamics, using measures from complex network theory.

9. Assignment Description:

3 Delivery

The deliverables for Project Assignment B are. A Paper (.pdf format). The paper should contain your analysis, it should tell the story about the data and the research finding. An Explainer Notebook (.ipynb format). The Notebook should contain all the behind the scenes stuff, You should link to the notebook from the paper (in the Methods section).

4 Paper Structure

reference videos:https://www.youtube.com/watch?v=vtppZ3wDVGs The paper should have all the elements that are in the templates (links below):

1. Abstract
2. Significance Statement
3. Introduction
4. Results
5. Discussion
6. Methods
7. References

Use the template's author contributions (it is not OK simply to write "All group members contributed equally".).BUT YOU ONLY HAVE MAX 5 PAGES(everything must be within those 5 pages, also the references) and max 5 figures. Less/shorter is OK, longer is not OK. The notebook should contain the code for your analyses. We appreciate if notebooks are commented and structured nicely. In addition, you can include the following information

5 grading

1. The main point is to show off what you've learned in the course, so the first thing is to make sure your dataset contains both networks and text.

2. That you did a thorough analysis that shows what you've learned in the class. (And we can only know about this if you use and show key parts of that analysis in the paper itself.)

3. Did you manage to get to a research finding about your dataset? (And not just reproduce the analyses from the lectures on your own dataset)

4. All the formal things in the paper-writing video. Good abstract, Intro, etc. Informative figures with thoughtful captions, etc. The right references. Readable text.

5. A well structured explainer notebook.

6 submitting

The notebook should contain the code for your analyses. We appreciate if notebooks are commented and structured nicely. In addition, you can include the following information

Data and Stats. Write in more detail about your choices in data cleaning and preprocessing Did you do analyses / calculate statistics that didn't make it to the main text, put them here. Handing in the assignment: Simply upload your .pdf to peergrade. And link to the Explainer Notebook on GitHub from the paper's method section. (we do check timestamps, so don't edit the Explainer Notebook after the handing date).

7 requirements from assignment

1. To help us navigate the Notebook, it's a good idea to repeat the question you're answering.

2. Try to control the length of your notebook. While grading, we look at how you prioritize material and express yourself clearly and succinctly.

3. Read the text carefully - make sure you understand the question. And make sure that you answer all sub-questions, etc. (It's easy to miss something, so be thorough).

4. Format your plots properly. Axes must be labeled, use

5. Make sure there is text explaining each figure. Use scientific papers as a gold standard for how to comment on figures.

6. Make sure that you use references when they're needed and follow academic standards.

7. Be precise, write in objective language (avoid: "I think ...", "In my opinon...", etc.) - if you make an observation, support it with data.

8. Answering "yes" or "no" to questions is not sufficient. The questions are an invitation for you to reflect on why.

9. Example of problematic response. Q: "Do the two arrays have the same length?" A: "no"

10. Guessing on or eyeballing the answer is not sufficient. In almost all situations, you should provide the exact solution.

11. Example of problematic response. Q: "calculate the number of nodes Ncr so that the network has only one component." A: "It can be seen from the plot that it is around value"

12. Show that you understand the context of the solution.

13. Make sure you comment your figures and explain which information you are able to extract from them even if it is not explicitly asked.

14. Make sure you add references any time you introduce a formula, algorithm, etc.

15. Make sure that your notebook does not contain irrelevant output

16. Example: do not submit a notebook, where you printed out a huge list of links in your graph