Department of Computer Science & Engineering

Aug-Dec, 2023

UE21CS343AB3: Graphs Theory and its Applications

Assignment 03

This is a team assignment and will have a team evaluation. You can refer to the sample code provided to you for this assignment.

English novels are available on the Project Gutenberg site. Choose one for your assignment.

Download the text version (not HTML), as you don't want to spend time in unnecessary pre-processing. Do the following :

1. **Implementation (In Jupyter Notebook):**
    a. Make a list of characters in the novel. You need to decide whom to include based on who matters. Do not choose a very big novel (Like War and Peace by Tolstoy or Mahabharata as there are too many characters) in it as pre-processing will be a challenge. Do not choose a novel that has only a minimal number of characters. Then, your analysis will not be interesting.
    b. Extract a social graph of the manually identified characters in the text. To do this, you need to use a co-occurrence algorithm (co-occurrence in a single sentence or every few sentences – you can decide based on trial and error. If a character has multiple names, you have to pre-process that into a single name.). Provide a Networkx Visualization of the graph (no need to use Gephi) .
    c. Calculate the four types of centrality of main protagonists, i.e. degree, betweenness, closeness, PageRank.
    d. Extract the Ego network and calculate the local clustering coefficient of the main protagonist nodes.
    e. Detect communities using the following methods: Clique Percolation Method, Girvan Newman, and Louvain algorithm (Modularity based)
    f. Suppose you have a social graph of 50 characters. Create equivalent generative models of 50 nodes to compare against the social graph that you extracted (Ref Generative models). You should try out Random Graph Model, Preferential Attachment and Small world model.

2. ***Analysis – Considering all your results, document your analysis point by point in the same Notebook in a separate cell, "ANALYSIS"***

   Theme of the analysis: What do you know of the story, and does it match with what you got from your network analysis? Have you got any insight to offer?

   a. Who are **the protagonists in the story, as per your analysis**? Interpret the **4 centralities.**
   b. **Comment on how "real world" the social graph is, i.e. is it based on fact or fiction?** Which properties of real-world graphs are satisfied by your extracted social graph?
   c. A story typically has dynamics. Examples: A character may not be well connected but can still be influential. Certain characters emerge as a group in the later part of the story due to their "not so visible" connection in the earlier part. **Does your analysis tell you about the dynamics in the story?**

Submission Instruction:

1. **Submit a single zip file of (a) Python Jupyter notebook (only one submission per team) (b) the pre-processed text file**. The TAs should be able to execute your notebook
   a. **The first cell** should have **SRN and NAME of your team members**.
   b. **The second cell** should point to the story file that you have used.
   c. **The third cell** should summarize the social graph extracted i.e. no of characters, no of edges, etc.
   d. **The fourth cell** should have all the required import statements for running your notebook
   e. The submitted file should be named using a convention of "**SRN1_SRN2_SRN3.zip**"
   f. Use Python 3.x and Networkx 2.5 or above
   g. Refer to the Networkx API reference while coding
2. **In the submitted Jupyter Notebook – make clear sections in the notebook as per the questions asked. The Analysis SHOULD BE AT THE END**
3. **Graph Visualizations are not asked for. Do not do it unless you think that you can convey a point by that. To do that, one can export Networkx Graph and Import it into Gephi.**
4. **Do not submit what is commonly available on Github. No marks will be given for such submissions.**