Assignment 3: Graph Data - Group Submission - Returned

Title	Assignment 3: Graph Data
Groups	Group B"
Students	Suseela Pattamatta (suseelapattamatta), Mona Malik (monamalik27), Nikita Kapoor (nikitakapoor), Sannath Vemula (sannathreddyv)
Submitted Date	Nov 7, 2019 11:21 PM
Grade	**4.00 (max 4.00)**

Instructions

Objective
By now, this is getting predictable. You are provided with some temporal graph data and asked to illustrate some exciting insights hidden within it. You have two weeks (until next Sunday), but you should demo a prototype visualisation to the whole class next Monday. There is no pre-defined "correct solution", but you need to demonstrate mastery of graph data modelling in the context of doing Data Science.

Data
Reddit is an American online community of message forums that is split into "sub-reddits" that each have a self-defined focus. You are provided with a *hyperlink graph* (https://snap.stanford.edu/data/soc-RedditHyperlinks.html), where each node represents one of the 55,000 sub-reddits and an edge indicates that a post in one sub-reddit linked to a post in another sub-reddit. The edges are directed *and* multivariate: they are labeled with a creation timestamp and a (crowd-sourced) *sentiment polarity* that indicates whether the linking content was explicitly negative towards the target sub-reddit. You can read more about the origins of the dataset on the project page of its creators: http://snap.stanford.edu/conflict/.

Be certain that you model graph-based aspects of the data. Relational and/or spatio-temporal aspects do not alone fulfill the expectations of the assignment.

Implementation
It is expected that you will use Python, as it is the language of choice in Data Science. Moreover, it provides many visualisation libraries that will help you achieve more than you realise you can! Also, by now you should be more comfortable with it.

Pre-submission Demo
In class on 28 Oct, each group will have 3-5 minutes to demo their prototype to the entire class, using a different presenter than the previous assignment. Please send a pdf with at most 4 pages by email to me prior to class so that we do not need to connect laptops to the projector. Also, a discussion will be taken collectively *after the last group presents*; so, a more exciting demo will likely garner more peer feedback.

Submission
You should submit your implementation in raw python or python notebook and a short technical report (5-10 pages) in pdf format. The markers will primarily grade the submission based on the report, but may validate that the implementation works as described. The report will describe:

- the insights revealed by the visualisation/application
- the design choices in the implementation/visualisation
- challenges encountered (if relevant)
- how this submission meets the requirements set out in the rubric (i.e., a justified self-evaluation)
- other details that you consider relevant

Grading
The entire group will receive a grade based on how well the submission adheres to the rubric below. Note that your report provides the opportunity to persuade the markers, but that they will ultimately grade according to the rubric.

The demo on 28 Oct will *not* form part of the grade; however, it is an opportunity to solicit peer and instructor feedback a few days before the deadline that may ultimately improve the quality of your submission (i.e., your grade). Also, having content to present will contribute to your participation grade in the course.

Assignment 2 Rubric

Component	Weight	4	3	2	1
Transforming Raw Data to Insights	40 %	Visualisation is very informative and tells a story about the data; visualisations are clear and easy to interpret without aide.	Visualisation is informative, albeit possibly with some support from text	Visualisation is complete; information is presented, but lacks complexity or depth.	Very difficult to extract any insights from the visualisation, or no visualisation at all
Data Modelling	30 %	Demonstrates a clear mastery of modelling graph data; modelling of the data is flawless and/or demonstrates knowledge of both the strengths and limitations of the data model.	Model is appropriate for the underlying dataset(s) and successfully incorporates the underlying complexity of the raw data.	Graph model exists, but is very simplistic or not used to produce final visualisation	Minimal use of concepts in graph data modelling

Algorithmic Considerations	20 %	Compelling evidence is provided to show the scalability of the submission, by design.	Algorithmic considerations are well reasoned and demonstrate that performance has been a design consideration.	Some attempt is made to address questions of efficiency, but the visualisation is restricted to small samples of the dataset.	Minimal attempt to design an efficient solution; visualisation perhaps does not load
Relationship to Graph Data Challenges	10 %	Research and/or industry challenges for graph-based data model are clearly addressed, adding substantial value to the project.	Research and/or industry challenges for graph-based data are considered, adding value to the project.	Some attempt is made to engage with research papers presented in class, but it adds limited value to the project	Does not engage with research papers presented in class nor other typical challenges faced with graph data

Tips

1) Try to load the data into a usable structure early (first few days). You may find it more difficult than you expect to model the raw datasets as workable data structures.

2) Ascend quickly to a working visualisation and add complexity later; i.e., first build a minimal viable product (MVP). You will be exposed to new research and new ideas in class as you work on the assignment, so you want to be *agile* with your development patterns. At the same time, even in groups, it may take longer than you expect to go from raw data to a working visualisation; so, you don't want to leave this until the last few days before the demo or you could end up with nothing to show at all!

3) Use libraries prolifically.

4) Consider addressing each rubric component individually in your tech report.

5) Reflect closely on the feedback for Assignments 1 and 2. This is meant to guide you towards a higher grade on subsequent assignments.

Additional resources for assignment

Original submission text with the instructor's comments inserted if applicable

Please find attached the report for assignment 3 for Team-B. Also attached are the plot diagrams and the Code files. Thank you

Submitted Attachments

- Insights Diagrams.zip (833 KB; Nov 7, 2019 11:21 pm)
- Code.zip (120 KB; Nov 7, 2019 10:49 pm)
- CSC 501 Assgn 3 report final.pdf (1 MB; Nov 7, 2019 11:20 pm)

Additional instructor's comments about your submission

Overall
4/4/4/4 = 4

Derivation of Insights
4
+ Extensive set of related visualisations around graph properties, e.g., degree, SCC's
- Could benefit from more discussion/interpretation of the visualisations

Data Modelling
4
+ Explicit comparison of several different data models, selected best performing by means of experimentation

Algorithmic Considerations
4
+ Explicit indication of relative run times
- Could benefit from *scalability* analysis; i.e., how does size of input affect time/memory?

Connection to Research
4
+ Took idea of chord diagram from graph viz paper, implemented it from scratch with js/html, using suggested packing algorithm. Great extension of course content to produce a beautiful graphic.