COSC 2671 Social Media and Network Analytics

Assignment 2

Choose Your Own Analysis

Learning outcomes:

- Build on analysis and problem solving skills developed in course and assignment 1 to use social media & network analysis to solve or answer data related questions/problems
- Practice collecting and exploring data
- Develop written and verbal communication of describing the problem, approach, analysis, insights and justification of approach taken.

Group size: 3

Due Date:

- Completed Assignment: 11.59pm Sunday, 20th May

Introduction

Previous assignment, you had an opportunity to practice doing analysis about topic modelling and sentiment analysis on Twitter. In this assignment, you'll build upon those skills and work in a team to solve/answer an interesting problem, analysis or question of your choice using social media and networks data.

Overview

Goals of the master programs and of RMIT graduates in general are the ability to problem solve, to work in teams and to be able to communicate. This assignment aims to contribute towards your development in these crucial areas. It is essentially a cross between a hackathon and a final year project, where you work on a problem to solve or answer in a team, and in a length of time that is between a hackathon and a final year project.

The assignment can be broken up into a few parts, corresponding to how you could approach it (and also corresponding with the marking rubric). The parts are as follows, and in the following we'll provide further details of each:

- Team formation
- Problem/Question construction
- Team management & Mentoring
- Perform the analysis/Solve the problem
- Communicating the results or analysis

Team Formation

This assignment is a team one, in groups of 3. This allows the scope to be larger so more interesting problems or analysis can be done, as well as allowing you to support each other in a team.
Initial task of the assignment is to form the teams. I can help you with this if you are having problems, but my strong suggestion is to select team members with similar approach to assignments and study. Ideally you also want a team with a good mix of skills, e.g., analytics, programming/data collection, communications etc. Of course, if everyone are strong in all areas that would be great.

Once you confirmed, go to Canvas, I have setup groups for assignment 2, where you can self-assign to one of the teams. Please only enter an allocation once you have the team formed.

Problem or Question Construction

You'll also need to select the problem or question(s) to work on. Here, I'll like your team to discuss, and then for each team to meet with me in person to ensure the problem/question is interesting (i.e., non-trivial) but doable in the 5 weeks you have. The assignment will not be marked if you haven't met me to discuss the problem by the 2nd week (end of April 27th). I can meet on Skype if need be.

To help you get started and to give you a rough idea of the scope of the problem that are doable, consider the following:

- Depression and suicide alert on social media
- Beyond statistics, who is the best and most influential cricketer?
- Who is the centre of the Marvel/Nintendo universe? Can we promote Browser to be the most central Nintendo head honcho?
- Github & communities – all things about community analysis on Github.
- Donald vs Kim – who has the bigger ego?
- Social Media Neighbourhood Watch – Alerts for where, when and what crimes are been committed in Melbourne?

If you look at the problems, some are more analysis/research in nature, others are about solving a problem. This is not an exhaustive list. There is no bias towards the different types, as I know each of you have different backgrounds and knowledge and interest, so would like to leave it open to you. There are some constraints that your assignment should adhere to though:

- Must include at least one source of social media and networks.
- Must include a graph or network and its analysis/processing to solve a problem.
- Must include some element of analysis using social media and networks, and something we learnt in class. It cannot be all machine learning for example.
- Your analyst or solution is in Python (no R, including visualisation), but parts maybe in other languages, e.g., Javascript if you decide to prototype a front end on website.
- All code should be your own, you can leverage packages such as networkx etc., but it shouldn't be copied from an existing solution.

If unsure please contact me.

Team Management & Mentoring

Although it isn't part of this course, I know from previous experiences as a student and lecturer about some of the difficulties in a team. Some of you are very experienced in working as a team, others may do with a few pointers, so some of the following is advice and suggestions that can help with your team dynamics and management:

- Each team is different and what works best is team dependent, but the most successful teams usually have either all highly motivated and self-disciplined members or a team leader who ensures the teams don't lose sight of targets and drive the team towards them. Regardless of style, I strongly suggest to nominate a leader, and to agree upon vesting some authority with them, e.g., keep everyone on track etc.
- Set up regular meetings within your team, whether in person or on Skype. Have an agenda for each meeting, and a plan for the assignment.
- Communicate with each other and try to stick to timelines/plans. If can't stick to them, have remedial actions.

The following are suggestions as well as requirements:

- Once you have someone as the team leader, please specify within the canvas groups. I'll mainly liaison with the leaders (from my end) and the leader can disseminate the information to the team members.
- Use timesheets and other means to record progress, which is part of your submissions and to avoid arguments later about contributions, and setup a code repository (e.g., Github or Bitbucket) and/or Google drive to easily share code. The repository/shared drive is a good way to keep everything together and also to ensure progress is made. In the event there is an argument about contributions, it is also a way for us to resolve who did what.

In addition to above, I'll like each team to meet me at least 3 times before the assignment is due. Once to discuss problem, once to discuss progress and your approach, and one in the final phrase of the assignment. Consider this as mentorship or guidance from me, as well as helping you with keeping on track. If you look at the marking rubric, there is marks associated with it.

I have set some appointment slots on Google calendar. If that doesn't work I'll setup a Doodle or some other approach for you to select a slot to meet.

Perform the analysis/Solve the problem

The next component is to gather the necessary data, proceed with your analyst and iterate. For this part, I encourage you to use any of the tools and techniques we have studied in class, but you are also most welcome to use techniques you have studied in other classes – e.g., data visualisation, time series, machine learning etc. The only restriction here is that all content in the report and presentation should be in Python or the analysis originated from Python and some visualisation tool was then used, it must use social media and networks, and it must use some elements of what we have studied in this course – i.e., text or NLP, social network analysis and behaviour analysis (see constraints about problem above). Do exploration of approaches, but I also want you to give consideration of why you are trying an approach and how it might help you solve the goals of your assignment.

Communication

Final component is to communicate your solution, analysis and discussions. This is done in two ways. One is via a report, much like what you did for assignment 1. The second is a group presentation, where I want each group to present and "pitch" their problem and analysis during the last week of the semester. As an added bonus, everyone will get to vote on what they think is the best presentation, and the highest one will get a (small) prize, bagging rights and if useful for you, I'll be happy to provide a letter indicating what you were judged for the best presentation and project.

Sources of Help

Your lecturer would be very happy to discuss questions and your results with you. Please feel free to come talk to me during consultation, or even a quick question, during lecture break. Use the existing communication channels you have been using with me also.

Also you can ask questions on canvas, but please do not post any code.

There will also be a FAQ, and anything in the FAQ will override what is specified in this specifications, if there is ambiguity.

Plagiarism

Remember to gain maximum benefit from our time together in this course, try things yourself. See https://www.rmit.edu.au/students/student-essentials/rights-and-responsibilities/academic-integrity for details about academic integrity.

This including asking or paying someone else to do your assignment. I am one of those lecturers that care about fair outcomes for all students.

Assessment

The assessment is based on your presentation, report and your progress.

Examine the assessment rubric below.

You'll be assessed on how well you communicated your work, on the problem selected, how your team went about the assignment and your insights/solution. As a rough rule of thumb, you can get a credit and distinction if you just used the code we learnt in class, but to get a HD would require you doing some research and going beyond what we did in class.

In addition, part of your assessment would be self-reflection on your own performance and your team's performance. Here you may add any additional information you'll like to inform me of difficulties that we didn't cover in the meetings. I'll also like you to provide an estimate of the work breakdown (via timesheets and your assessment of contributions in the self-reflection), and to ensure as even work distribution as possible, it will up to my discretion for marks redistribution if the workload is unreasonably unbalanced.

What to submit

- Your report, up to 30 A4 pages in length and in font size 12 of assessed content, not including appendices. Note this is a maximum, not a length you need to must have. Anything beyond 30 pages will not be read, and anything in appendix should be considered as additional information, as there is no guarantee it will be read. Please do not ask for more pages, this limit is strict.
- Your scripts used to perform your analysis. Please comment and style the code, as we will read them. Note that I may also ask you to explain the code and why you chose particular coding choices.
- Share access with me to your code and/or document shared repository or drive. Include the details how I can access it.
- Timesheets, project plans, individual self-reflection (max of one A4 page)
- A sample of the data, no more than 10 Mb in size.

Submission should be made to Canvas. Closer to submission we will describe the submission process in more detail.

Presentation

More details about the presentation/pitch will be available when we get closer to it, but essentially your team should tell the class about the (exciting) problem or question your team tackled, who you think it is an important/useful/interesting/exciting, your approach and your results, solution or insight. There should be time allocated for questions also.

Rubric

Use the following rubric to help you determine how to approach the assignment.
Note teamwork and management is an individual mark, based on how your team functions, manages itself and your self-reflections.

Criteria	Excellent	Good	Fair	Poor
Problem Formulation (15%)	The proposed problem/question is interesting, well-motivated and non-trivial. Problem and success criteria are well designed and thought out, and scope is realistic.			
Teamwork and Management (10%)	Team functions well and have effective management procedures. Have a well-designed project plan, weekly plans. Timesheets are			

Criteria	Excellent	Good	Fair	Poor
	realistically filled in and individual self-reflection is insightful and reflects thought has been given to what worked, what didn't and how to improve in the future. Met with Jeff at least 3 times.			
Approach (20%)	The approach is an appropriate method to take to solve the problem or answer the analytical question. Approach taken goes beyond using the tools provided in class. Team justifies and explains their approach well. Approach includes data collected and techniques used.			
Analysis/result & Discussion (25%)	Problem solving: The solution solves the problem well and all the success criteria are satisfied. Team is able to provide analytical and/or empirical evidence of this. Answering analytical question & Analysis component: Excellent discussion of results that answers the question proposed or contributes			

Criteria	Excellent	Good	Fair	Poor
	towards solving the problem. Conclusions are supported by analytical and/or empirical evidence. All success criteria are answered.			
Report Presentation (15%)	Report is easy to read and flows well. It is structured well, leading the reader through the process of answering the questions or solving the problem. Tables, figures and other visualisation are easy to read and to interpret.			
Group presentation (25%)	Presentation is well structured, clear and easy to follow by master students. It is engaging and interesting and able to capture the audience's attention. A candidate for best presentation.			