
Social Media and Network Analysis
COSC 2671
Assignment 2

Assessment Type	Group assignment. Submit online via Canvas → Assignments → Assignment 2. Marks awarded for meeting requirements as closely as possible. Clarifications/updates may be made via announcements/relevant discussion forums.
Due Date	May 30, 11:59pm
Marks	40

1 Overview

In the previous assignment, you had an opportunity to practice using topic modelling and sentiment analysis to perform analysis on Reddit. 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 data.

2 Learning Outcomes

This assessment relates to the learning outcomes of the course which are:

- CLO 1: Apply data science to analyse social media and social networks;
- CLO 2: Analyse social networks by finding communities, identifying important nodes, and influence propagation;
- CLO 3: Analyse social media by applying Natural Language Processing (NLP) techniques to detect sentiment and topics;
- CLO 5: Synthesise and present insights from the social media and network analysis performed.

3 Assignment Details

The 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

3.1 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. We can help you with this if you are having problems, but our 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 have confirmed your team, go to Canvas, We 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, and if your team has less than 3 members, please kindly contact us to explain your reasoning about this. Please do not put yourself in a team in the hope others will join you – use the appropriate discussion forum for that first to form a team.

3.2 Problem or Question Construction

You'll also need to select the problem or question(s) to work on. Here, we'll like your team to discuss, and then for each team to meet with us in person or via Teams to ensure that the problem/question is interesting (i.e., non-trivial) but doable in the time you have.

As a guide about the scope of problems that are doable, consider the following:

- 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?
- What is trending in the area of Climate change on social media?
- What are the characteristics of successful and popular designs on Canva?

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 in assessment towards the different types, as we 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.
- 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 analysis 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 us.

3.3 Team Management & Mentoring

Although it isn't part of this course, we know from our previous experiences as a student and lecturer that there are potential difficulties when working 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, we 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, on MS Teams, Skype or Slack. 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.
- Once you have someone as the team leader, please specify within the canvas groups. We'll mainly liaison with the leaders (from our 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), a MS Teams team 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 the above, we'll like each team to meet us every week before the assignment is due. To discuss the proposed problem, approach and progress, and final wrapping up of the project. Consider this as mentorship or guidance from us, as well as helping you with keeping on track. If you look at the marking rubric, there are marks associated with this.

We will set some appointment slots on a calendar (Outlook doesn't have this natively). If that doesn't work we'll setup a Doodle or some other approach for you to select a slot to meet.

3.4 Perform the analysis/Solve the problem

The next component is to gather the necessary data, proceed with your analysis and iterate. For this part, we 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 or community detection (see constraints about problem above). Do exploration of approaches, but we 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.

3.5 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 we want each group to present and “pitch” their problem and analysis during the last week of the semester. Each member must speak in the presentation.

4 Submission

- Your report, up to 40 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 40 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 we may also ask you to explain the code and why you chose particular coding choices.
- Share access with us to your code and/or document shared repository or drive. Include the details how we 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.

5 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 us of difficulties that we didn't cover in the meetings. We'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 if possible, it will up to our discretion for marks redistribution if the workload is unreasonably unbalanced.

6 Academic integrity and plagiarism (standard warning)

Academic integrity is about honest presentation of your academic work. It means acknowledging the work of others while developing your own insights, knowledge and ideas. You should take extreme care that you have:

- Acknowledged words, data, diagrams, models, frameworks and/or ideas of others you have quoted (i.e. directly copied), summarised, paraphrased, discussed or mentioned in your assessment through the appropriate referencing methods
- Provided a reference list of the publication details so your reader can locate the source if necessary. This includes material taken from Internet sites. If you do not acknowledge the sources of your material, you may be accused of plagiarism because you have passed off the work and ideas of another person without appropriate referencing, as if they were your own.

RMIT University treats plagiarism as a very serious offence constituting misconduct. Plagiarism covers a variety of inappropriate behaviours, including:

- Failure to properly document a source
- Copyright material from the internet or databases
- Collusion between students

For further information on our policies and procedures, please refer to the following: <https://www.rmit.edu.au/students/student-essentials/rights-and-responsibilities/academic-integrity>.

We will run both code and report similarity checks.

7 Getting Help

There are multiple venues to get help. First point of call should be Canvas, assignment FAQ (on Canvas), recordings about it and the discussion forum. There will also be weekly online consultation hours for the assignment (see Canvas for times). Please **refrain from posting solutions** to the discussion forum.

A Rubric (40 marks)

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

Criteria	Excellent	Good	Fair	Poor
Problem Formulation (5 marks)	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.	The proposed problem/question is interesting, relatively well-motivated and non-trivial. Problem and success criteria are reasonably designed and thought out, and scope is realistic.	The proposed problem/question is somewhat interesting, has some motivation and on the simple/trivial side. Problem formulated for the sake of completing the assignment. Problem and success criteria are not well designed and thought out, and scope may be unrealistic.	The proposed problem/question is not very interesting, has minimal motivation and on the simple/trivial side. Problem formulated for the sake of completing the assignment. Problem and success criteria are not well designed and thought out, and scope may be unrealistic.
Teamwork and Management (including weekly catchup) (7.5 marks)	Team functions well and have effective management procedures. Have a well-designed project plan, weekly plans. Timesheets are 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 the lecturer at least 3 times.	Team functions well and have reasonable management procedures. Have a decent project plan, weekly plans. Timesheets are realistically filled in and individual self-reflection is reasonably insightful and reflects thought has been given to what worked, what didn't and how to improve in the future. Met with the lecturer at least 3 times.	Team functions more as a collective of individuals and there some communications demonstrated between members. Management procedures are somewhat lacking. Have a project plan and weekly plans that reads to be done before submission. Timesheets appear manufactured and individual self-reflection is not very insightful and rushed. Met with the lecturer at least 3 times.	Team functions more as a collective of individuals and there little communications demonstrated between members. Management procedures are lacking. Lacking or minimal project plan and weekly plans that reads to be done before submission. Timesheets appear manufactured and individual self-reflection is not insightful and rushed. If team did not meet Imesh/Lida at all, this is the maximum mark one can expect for this criteria, regardless of other aspects submitted.
Approach (7.5 marks)	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.	The approach is an appropriate method to take to solve the problem or answer the analytical question. Approach taken is limited to mostly the tools provided in class. Team justifies and explains their approach well. Approach includes data collected and techniques used.	There are other approaches that are clearly more appropriate method to take to solve the problem or answer the analytical question. Approach taken is limited to the tools provided in class. Team somewhat justifies and explains their approach, but there are unexplained choices. Approach includes data collected and techniques used.	There are other approaches that are clearly more appropriate method to take to solve the problem or answer the analytical question. Approach taken is limited to the tools provided in class. Team does not justifies and explains their approach. Approach includes data collected and techniques used.
Analysis/result & Discussion (10 marks)	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 towards solving the problem. Conclusions are supported by analytical and/or empirical evidence. All success criteria are answered.	Problem solving: The solution solves the problem adequately and most of the success criteria are satisfied. Team is able to provide analytical and/or empirical evidence of this. Answering analytical question & Analysis component: Good discussion of results that answers the question proposed or contributes towards solving the problem. Conclusions are supported by analytical and/or empirical evidence. Most success criteria are answered.	Problem solving: The solution somewhat solves the problem and about half of the success criteria are satisfied. Team provides some analytical and/or empirical evidence of this. Answering analytical question & Analysis component: Adequate discussion of results that answers the question proposed or contributes towards solving the problem. Conclusions are somewhat supported by analytical and/or empirical evidence. About half the success criteria are answered.	Problem solving: The solution does not really solve the problem and many of the success criteria are not satisfied. Team is unable to provide analytical and/or empirical evidence of this. Answering analytical question & Analysis component: Minimal discussion of results that answers the question proposed or contributes towards solving the problem. Conclusions are not supported by analytical and/or empirical evidence. Many of the success criteria are not answered.
Report Presentation (5 marks)	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.	Report is reasonably easy to read and flows relatively well. It is structured reasonably 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.	Report is difficult to follow in places and doesn't flows well. It is adequately structured, but reader may find it difficult to follow through the process of answering the questions or solving the problem. Tables, figures and other visualisation are either too small or difficult to interpret.	Report is difficult to follow and doesn't flows well. It is barely structured, but reader may find it difficult to follow through the process of answering the questions or solving the problem. Tables, figures and other visualisation are either too small or difficult to interpret.
Group Presentation (5 marks)	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.	Presentation is reasonably well structured, clear and easy to follow by master students. It is reasonably engaging and interesting and able to capture the audience's attention.	Presentation has some structure, somewhat clear and easy to follow by master students. It is not that engaging and interesting and unable to capture the audience's attention.	Presentation has little structure, is not clear and not easy to follow by master students. It makes little attempt to be engaging and interesting and unable to capture the audience's attention.